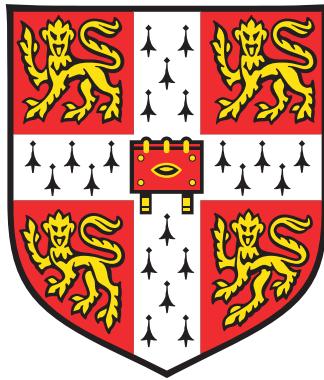


Rare events and dynamics in non-equilibrium systems



Lukas Takaaki Kikuchi

Department of Applied Mathematics and Theoretical Physics
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Declaration

I hereby declare that except where specific reference is made the work of others, the contents of this dissertation are original, and includes nothing which is the outcome of work done in collaboration except as indicated in the text. It is not substantially the same as any that I have submitted, or am concurrently submitting, for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted, for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

Lukas Takaaki Kikuchi

October, 2022

To my friends.

Acknowledgements

Abstract

Rare events and dynamics in non-equilibrium systems

Lukas Takaaki Kikuchi

The matter of this thesis is divided in two parts, both of which belong to disciplines that lie within the purview soft matter physics. In the first part, we study the infinite-dimensional probability space of stochastic differential equations. In particular, we study the ensemble of transition paths between meta-stable states of Itô diffusions. We develop a suite of techniques to characterise the dominant transition channels of such systems, and study the character of the transition path ensemble as a function of varying diffusivity.

rare events

Contents

Introduction	13
I Global methods for sampling rare diffusive events	19
1 Ritz methods for Freidlin-Wentzel-Graham actions	21
1.1 Introduction	21
1.2 Large deviation theory	23
1.3 On-shell action	26
1.4 Ritz method	29
1.5 Numerical results	33
1.5.1 Brownian dynamics in a complex potential	34
1.5.2 Brownian dynamics in a circulatory field	36
1.5.3 Multistability in a genetic switch system	37
1.5.4 Transitions between limit cycles	38
1.5.5 Egger model of weather	40
1.6 Numerical convergence	41
1.7 Conclusion	41
2 Monte Carlo methods on path spaces	45
2.1 Introduction	45
2.2 Path-measures and densities for stochastic processes	47
2.3 Path-space MCMC methods	53
2.3.1 The preconditioned Crank-Nicolson algorithm	54
2.3.2 Kosambi-Karhunen-Loève expansions of stochastic trajectories . . .	59
2.3.3 Numerical algorithm	61
2.3.4 Mode-space band structure	64
2.3.5 Mode-dependent step-sizes	70
2.3.6 Generalised end-point conditions	72
2.4 Sampling multi-modal transition path ensembles	73
2.4.1 Second-order variational expansions of stochastic action functionals	76

2.4.2	The Teleporter MCMC method	78
2.4.3	Numerical algorithm	82
2.5	Conclusion	85
3	Diffusivity dependence of transition paths	89
3.1	Introduction	89
3.2	The transition path ensemble	90
3.2.1	Model system	90
3.3	Semi-analytical approximations of stochastic path measures	91
3.3.1	The Gaussian mixture approximation	92
3.3.2	Approximations of transition channel probabilities	93
3.4	Results	94
3.5	Conclusion	97
II	Geometric mechanics of microstructured soft matter	99
4	Introduction	101
4.1	Classical Cosserat theory	102
4.2	Mathematical preliminaries	107
4.2.1	Differential geometry	108
4.2.2	Exterior calculus	109
4.2.3	Lie groups	112
4.2.4	Homogeneous spaces	115
4.2.5	The Maurer-Cartan form	116
4.2.6	Sub-manifolds of homogeneous spaces	117
4.2.7	Vector and matrix operations	120
5	Cosserat rod models and geometric mechanics	123
5.1	Geometric mechanics of a rigid body	124
5.2	Geometric Cosserat rod mechanics	129
5.2.1	Cosserat rod kinematics	129
5.2.2	Cosserat rod dynamics	137
5.2.3	Summary and discussion	146
5.3	Applications	147
5.3.1	Hyperelastic slender rod	148
5.3.2	Dissipative and overdamped dynamics	151
5.3.3	Overdamped and inextensible Cosserat rods	153
5.4	Mechanics of kinematically constrained Cosserat rods	154
5.4.1	Kinematic constraints and gauge freedoms	154

5.4.2	Kinematically constrained dynamics	159
5.5	Conclusion	162
6	Geometric continuum mechanics on Lie group- and homogeneous configuration spaces	165
6.1	Generalised geometric Cosserat theory	166
6.1.1	Kinematics	167
6.1.2	Dynamics	170
6.1.3	Kinematic constraints	172
6.1.4	Summary and discussion	174
6.2	Derivation of the geometric kinodynamical equations of motion	178
6.2.1	Kinematics	178
6.2.2	Dynamics	181
6.3	Applications	187
6.3.1	Three-dimensional Cosserat media	188
6.3.2	Cosserat surfaces	191
6.3.3	Cosserat rods on 2-spheres	194
6.3.4	Relativistic Cosserat rods	197
6.3.5	$O(n)$ -NLSM field theory	201
6.4	Conclusion	202
7	Geometric numerical integrators	205
7.1	Geometric integrators for general configuration spaces	207
7.2	Geometric integrators for $SO(3)$ -configured system	210
7.3	Geometric integrators for $SE(3)$ -configured systems	211
7.4	Benchmark tests	213
7.5	Conclusion	216
Summary		219
References		223
Appendices		249
A	MCMC benchmark models	249
A.1	The asymmetric double-well system	249
A.2	The switch system	250
B	Calculating the regularised normalisation constants of infinite-dimensional Gaussian measures	253

C Reconstructing the spatio-temporal configuration	257
C.1 Numerical algorithm for reconstructing the Cosserat rod	257
D Detailed derivation of the $SE(3)$ short-time propagator	259
D.1 Simplified integrator for $SE(3)$ -configured systems	261
E Non-dimensionalisation of the Cosserat rod equations of motion	263
F Details of the geometric integrator benchmark simulations	267
F.1 Simulation of the Cosserat rod	267
F.2 Simulation of the Cosserat surface	272

Introduction

This thesis explores two distinct topics. The first part concerns itself with the study of transition events in stochastic differential equations, both in regimes where these are ubiquitous as well as rare. The second part develops a geometrical theory of Cosserat materials [1], and further a generalised framework for studying continuum systems with configurations taking value in Lie groups or homogeneous spaces. Whilst substantially different, the two parts of this thesis do share a common trait in that they explore the geometry of spaces that possess an infinite-dimensional character. In essence, the research we present in the first part of the thesis is to quantitatively and qualitatively investigate the concentration of path-probability measures, defined over the infinite-dimensional space of stochastic transition paths. In the second part of this thesis, we consider systems with infinite degrees-of-freedom; that is, the systems are Lie group-valued *manifolds* and we formulate their mechanics in terms of their intrinsic geometry. Below we will separately introduce the two parts of this thesis.

Introduction to Part I

Bolhuis: Studying transitions between stable states separated by unknown dynamical bottlenecks presents well known computational difficulties. Such transitions are rare events. Conventional methods for studying rare events in many body systems are based on transition state theory. First, one estimates the probability of visiting the dynamical bottleneckthe transition state $s!$. Then the transmission coefficient and rate constant can be determined from a series of short trajectories initiated from the ensemble of states prepared at that bottleneck.^{1–4} The difficulty with this method is that it requires prior knowledge of the transition states. Without knowledge of the transition states, i.e., without an a priori conception of the mechanism or pathway for the process, this approach cannot be applied.

Bolhuis 2: DISPARITY OF TIMESCALES Often, dynamical processes of interest occur on timescales that are very long compared to the shortest significant timescale. For example, the dissociation of a weak acid in water might occur with a half-life of, say, 1 ms, while elementary steps of molecular motions in water occur in 1 fs. Similarly, timescales for

folding the smallest of proteins are in the range of microseconds to milliseconds, whereas that for small-amplitude motions of amino acid side chains and water solvent is again 1 fs. This wide disparity of timescales can present serious computational challenges. For instance, consider a computed trajectory for a system containing a weak acid molecule and a bath of a few hundred water molecules. Within one or two orders of magnitude—depending on computing equipment and algorithm—1 s of computation time is required to advance the system for what would correspond to 1 fs of physical time. As such, typically 1012 s of computing time seems to be required to find one example of an event leading to acid dissociation. A representative sampling of pathways to dissociation would therefore seem to be an impractical computational task.

IMPORTANCE SAMPLING The basic idea is a generalization of standard Monte Carlo procedures (20, 21, 24, 25) that focuses upon chains of states constituting dynamical trajectories (26) rather than upon individual states. In its standard form, a Monte Carlo calculation performs a random walk in configuration space. The walk is biased to ensure that the most important regions of configuration space are adequately sampled. Specifically, in a Monte Carlo random walk, configuration x is visited in proportion to its probability $p(x)$. The walk may be initiated far from a typical configuration [i.e., x far from values of x where the weight from $p(x)$ is largest], but after some equilibration period, the bias drives the system to those important regions of configuration space. This feature is crucial to the success of Monte Carlo sampling. It is called importance sampling and is illustrated in Figure 2. Importance sampling can be generalized to trajectory space, as we have done to create the methods of transition path sampling. Consider the ensemble of all trajectories that are, say, 1 ps long. Most of these trajectories will be localized near some basin of attraction—a long-lived collection of neighboring microstates. Rare transition state crossings will comprise a small subset of these 1-ps trajectories. For example, if the process of interest occurs roughly once every millisecond, then only one out of a billion 1-ps trajectories will exemplify that process. Transition path sampling provides an efficient means to sample such rare subensembles.

Bolhuis 3:

The kinetics of many important molecular processes, such as phase transitions, self-assembly, conformational changes, association/dissociation, and chemical reactions, are dominated by events that are rare on the molecular time scales.[1] Straightforward simulations of such events are very ineffective, due to the long timescales involved, which are often related to high free energy barriers between metastable states. In the past fifty years many simulation methods have been developed that enhance the sampling of the unlikely regions of the barrier (e.g., see refs. [2–9]) or enhance transitions between states (e.g., see refs. [10–12]). However, the effectiveness of this enhanced sampling often depends on choice of a good reaction coordinate. Failure to capture the correct reaction coordinate leads to wrong

estimates of the barrier height, rate constants and mechanistic details.[13] Path-based approaches circumvent these problems by defining a dynamical trajectory connecting reactant and products states. Such paths can then be optimized to yield a minimum (free) energy pathway between the states. Well known examples of this approach include Nudged Elastic Band,[14–16] string method,[17,18] action minimization,[19] and dominant reaction pathway[20] techniques. For rough energy landscapes a more fruitful approach might be to sample from a distribution of dynamical paths. Transition path sampling (TPS) creates an ensemble of unbiased dynamical trajectories between the reactant and products states, through performing a random walk in trajectory space via a Metropolis Monte Carlo procedure.[

Grafke:

Small random perturbations often have a lasting effect on the long-time evolution of dynamical systems. For example, they give rise to transitions between otherwise stable equilibria, a phenomenon referred to as metastability that is observed in a wide variety of contexts, e.g. phase separation, population dynamics, chemical reactions, climate regimes, neuroscience, or fluid dynamics. Since the time-scale over which these transition events occurs is typically exponentially large in some control parameter (for example the noise amplitude), a brute-force simulation approach to compute these events quickly becomes infeasible. Fortunately, it is possible to exploit the fact that the mechanism of these transitions is often predictable when the random perturbations have small amplitude: with high probability the transitions occur by their path of maximum likelihood (PML), and knowledge of this PML also permits to estimate their rate. This is the essence of large deviation theory (LDT) [20], which applies in a wide variety of contexts. For example, systems whose evolution is governed by a stochastic (ordinary or partial) differential equation driven by a small noise or by a Markov jump process in which jumps occur often but lead to small changes of the system state, or slow/fast systems in which the fast variables are randomly driven and the slow ones feel these perturbations through the

TPS introduction.

FW -i "On the other hand, geometrical methods have revolutionized the field of classical mechanics since its inception, yielding some of the most important analytical techniques available in modern physics. These include Lie's theory of dimensional reduction through symmetries, and the alternative formulation"

However, cold limit not always physical. Ch 3 diffusivity dependenc ec

Ch by Ch breakdown.

In the limit of zero diffusion. The theory of Freidlin, Wentzell and Graham [2, 3, 4] gives asymptotic probability estimates of rare events in dynamical systems perturbed by small noise [5, 6, 7, 8]. Specifically, Freidlin-Wentzell-Graham (FWG) theory yields estimates of the stationary distributions and mean first-passage times. Both these quantities are

determined, in turn, by the asymptotic estimate of the probability of a stochastic trajectory to not deviate from a smooth path by more than a given amount in a given interval of time. The key result of FWG is that the limiting form of this probability, for small noise and small deviations, is given by a non-negative functional of the smooth path. This functional is known as the *Freidlin-Wentzell action* and its minimum, for fixed initial and terminal states, determines both the stationary distributions and first-passage times. The smooth path minimising the action is often called the Freidlin-Wentzell instanton. The theory is applicable to dynamical systems of both gradient and non-gradient character and can so be used to study a wide variety of equilibrium and non-equilibrium systems modelled by Itô diffusions [9, 10, 11, 12, 13, 14, 15, 16, 17, 18].

The set of realisations of a stochastic process is a functional space, and probability measures over such spaces are therefore infinite-dimensional. The research we present in the first part of the thesis

These are *global* in the sense of

In the first part of this thesis we study and develop methods for the study of transition paths

In Ch. 1 we began our by considering the transition path ensemble in the so-called Freidlin-Wentzell limit, which corresponds to a limit of vanishing diffusivity. **In these regimes transition events between metastable states are rare.**

In the first part of this thesis, we study and develop tools for the study of the transition path ensemble. Beginning at a limit of zero temperature, in Ch.

The fluctuating dynamics of many physical, chemical and biological systems are commonly modelled by stochastic differential equations expressed in Langevin or Itô forms [19, 20, 21, 22]. In such systems it is often of great interest to identify the typical pathways that stochastic paths take to transition from an initial to a final state, as for example in the nucleation of solids, the conformational changes in biomolecules, or shifts in ecological balance [9, 10, 13, 14, 15, 16, 17, 23, 24]. Typically, such transition paths cluster around multiple pathways in the space of configurations and the relative probability of one or the other of these pathways depends on the drift, the diffusivity, and the duration allowed for the transition to take place [25, 26, 27, 28]. As transitions are often rare events, direct simulations are not always practical and other means, analytical or numerical, are required to study them. Methods that allow for a full exploration of the space of transition pathways in stochastic dynamical systems, then, are of substantial theoretical and practical importance.

The theory of large deviations [2, 29, 30, 31], the topic of Ch. 1, provides an analytical method for obtaining transition pathways - instantons - in regimes dominated by the drift and for very long durations of path. Experimental systems, however, are typically not in a regime where the diffusivity is asymptotically low and durations are asymptotically long

[32]. While the relevance of including finite-temperature fluctuations around the instanton [33] is increasingly recognized [34, 35, 36], the physical implications of these fluctuations are far from being understood.

In this chapter, we show that the competition between drift and diffusion in transition pathways can be studied using the semi-classical expansions of the path measure of the stochastic dynamics introduced in Sec. 2.4. We use a mixture of Gaussian measures to approximate the path measure around its local instantons. This allows us to demarcate and transcend the boundaries of the low diffusivity regime, without recourse to sampling the TPE directly using MCMC methods. We demonstrate this explicitly for a two-dimensional overdamped mechanical system, with both conservative and non-conservative forces. For this system we uncover a counterintuitive phenomenon where typical transition paths do not concentrate around the most probable path, even at low-to-intermediate diffusivities where the Gaussian approximation is still valid. To validate our results numerically, and to study the TPE at high-diffusivity regimes, we use the teleporter MCMC (TMC) method developed in Ch. 2, that allows for simultaneous exploration of multiple transition pathways. We find excellent agreement between the semi-classical expansion and numerical results for a large range of diffusivities and path durations. The results presented in this chapter was the result of collaborative work published as a pre-print in [37], of which the author of this text was the primary contributor.

Introduction to Part II

In geometry part, you can talk about how intro, cosserat and geometry chapters serve as three successive dry runs of the same thing.

”On the other hand, geometrical methods have revolutionized the field of classical mechanics since its inception, yielding some of the most important analytical techniques available in modern physics. These include Lie’s theory of dimensional reduction through symmetries, and the alternative formulation”

Many systems in noneq where you have diffusions on lie groups.

Mention what the difficulties are with the conventional methods.

Maybe incorporate stuff from your talk in the geometry part of the intro. triangulation coordinateitisation. It is general difficult to triangulate a principal fibre bundle.

A wide range of structures exhibit properties of softness and slenderness, from biological materials like cilia [38] and cellular membranes [39], to strands of DNA [40] and hydrogels [41]. In this instance ‘soft’ refers to the constitutive properties of the material, like that of rubber or biological tissue, and ‘slender’ refers to its geometrical thinness.

- Also include references to research that have been done using cosserat theory.

We develop a general geometric theory of the kinematics and dynamics of G or G/H -

valued systems. We show in particular that the continuum mechanics of all directed media, otherwise known as Cosserat media, can be unified in description using this framework. As a special case, we also recover classical (undirected) continuum mechanics, expressed in Cartan's theory of moving frames.

Make it clear what is new: The general geometric framework and the Lie group integrators, as well as the presentation of all of the various Cosserat theories through the same framework.

Whilst the connection between Lie groups and mechanics has long been studied [42], here we provide a theory of mechanics Lie group-valued continua.

At a level of abstraction sufficient to describe all Cosserat theories, but also more general systems. For example, the framework is amenable to relativistic systems as well.

A lot of the results have already been shown before (a lot of the stuff in the Cosserat rod section in particular). The main purpose is to show how it all falls out of the a unified framework, and also works as a review article to some extent.

In this thesis, we

In the second part of this thesis

Mention that the two parts can be combined by deriving stochastic dynamics for part 2 systems and the field stuff for part 1.

in all the subsection conclusions, you will talk about further work. here you can summarise the further work.

Part I

Global methods for sampling rare diffusive events

Chapter 1

Ritz methods for Freidlin-Wentzel-Graham actions

1.1 Introduction

The theory of Freidlin, Wentzell and Graham [2, 3, 4] gives asymptotic probability estimates of rare events in dynamical systems perturbed by small noise [5, 6, 7, 8]. Specifically, Freidlin-Wentzell-Graham (FWG) theory yields estimates of the stationary distributions and mean first-passage times. Both these quantities are determined, in turn, by the asymptotic estimate of the probability of a stochastic trajectory to not deviate from a smooth path by more than a given amount in a given interval of time. The key result of FWG is that the limiting form of this probability, for small noise and small deviations, is given by a non-negative functional of the smooth path. This functional is known as the *Freidlin-Wentzell action* and its minimum, for fixed initial and terminal states, determines both the stationary distributions and first-passage times. The smooth path minimising the action is often called the Freidlin-Wentzell instanton. The theory is applicable to dynamical systems of both gradient and non-gradient character and can so be used to study a wide variety of equilibrium and non-equilibrium systems modelled by Itô diffusions [9, 10, 11, 12, 13, 14, 15, 16, 17, 18].

Determining the minimum of the Freidlin-Wentzell action is a problem in the calculus of variations. The Euler-Lagrange equations provide the necessary conditions for extrema of variational problems and, unsurprisingly, have been the basis of the large literature devoted to the numerical computation of Freidlin-Wentzell instantons [9, 43, 44, 45]. There exists, however, an alternative “direct” route for the solution of variational problems in which the functional is reduced, through finite-dimensional parametrisations of paths, to a multivariate function and then extremised by appropriate multivariate optimisation methods [46, 47]. To the best of our knowledge, the first use of the direct method for the Freidlin-Wentzell action, discretised by finite-differences, appears in the work of Weinan,

Ren and Vanden-Eijnden [48].

Here we combine the direct method with a Ritz discretisation [46, 47, 49] to minimize the Freidlin-Wentzell action. We analyse paths in a spectral basis of Chebyshev polynomials and use spectral quadrature to express the action as a multivariate function of the basis coefficients. Nonlinear optimisation is used to obtain coefficients that give the least action from which the instanton is synthesised in the spectral basis. For minimisation over paths regardless of their duration, this procedure is especially effective when applied to a reparametrisation-invariant on-shell form of the action that follows from the time-translational invariance of the Lagrangian. This generalises the scalar work functional of Olender and Elber (for gradient dynamics) and the geometric action of Heyman and Vanden-Eijnden [44] (for non-gradient dynamics). Our method is efficient enough to robustly sample the logarithm of the asymptotic estimate of the stationary distribution, *i.e.* the quasipotential, avoiding the alternative, but numerically delicate, route of solving the Hamilton-Jacobi equation [50, 51, 52]. Our method is simple to use, converges rapidly, and is applicable to both equilibrium and non-equilibrium problems. Its implementation is freely available on GitHub as the open-source Python library PyRitz [53]. The results presented in this chapter was the result of collaborative work published in [54], of which the author of this text was the primary contributor.

The remainder of this chapter is organized as follows. In the next section, we recall key results of Freidlin-Wentzell theory from dual perspectives of Itô stochastic differential equations and the corresponding Fokker-Planck equations. Sec. 1.3 we present the derivation of the on-shell form of the Freidlin-Wentzell action in a manner reminiscent of the Routh reduction procedure in classical mechanics and explain its relation to the scalar work and the geometric action. In Sec. 1.4 we describe the direct method for the minimisation of functionals, the Chebyshev spectral basis in which we construct smooth paths, the spectral quadrature rule we use to evaluate the action, and the multivariate non-linear optimisation methods we employ to find the minimum. In Sec. 1.5 we apply the direct method to three well-known diffusion processes and demonstrate convergence in each case.

A particular achievement of our approach is its relatively facile ability to calculate quasipotentials. This can be done with a sufficiently high density of sample points to construct effectively continuous maps of the quasipotential, which we do here for the same set of benchmark problems. We conclude with a discussion on extending the method to degenerate diffusion processes, systems with inertia and to the stochastic dynamics of fields.

1.2 Large deviation theory

We consider the autonomous dynamics of a d -dimensional coordinate $X = (X^1, \dots, X^d)$ in \mathbb{R}^d perturbed by configuration-dependent noise of intensity $\sqrt{\varepsilon}$ described by the Itô diffusion equation [55, 56]

$$dX^\mu = a^\mu(X)dt + \sqrt{\varepsilon}\sigma_\nu^\mu(X)dW^\nu \quad (1.1)$$

governing the stochastic trajectory $X(t)$, where $a^\mu(X)$ is the drift vector, $\sqrt{\varepsilon}\sigma_\nu^\mu(X)$ is the volatility, $W^\nu(t)$ is a d -dimensional Wiener process and repeated indices are summed over. Eq. 1.1 is also often referred to as an *overdamped Langevin equation*.

The transition probability density of the process, $P_{1|1}(x, t|x_0) = P(X(t) = x|X(0) = x_0)$, obeys the Fokker-Planck equation $\partial_t P(x|x_0) = \mathcal{L}P(x|x_0)$ where the Fokker-Planck operator is

$$\mathcal{L}(x) = -\frac{\partial}{\partial x^\mu}a^\mu(x) + \frac{\varepsilon}{2}\frac{\partial^2}{\partial x^\mu\partial x^\nu}b^{\mu\nu}(x) \quad (1.2)$$

and $b^{\mu\nu}(x) = \sigma_\lambda^\mu(x)\sigma_\sigma^\nu(x)\delta^{\lambda\sigma}$ is the diffusion tensor. We assume it to be non-degenerate, positive-definite and invertible. The inverse, $b_{\mu\nu}(x)$, induces a Riemannian structure in \mathbb{R}^d with a norm $|x|_b = \sqrt{b_{\mu\nu}x^\mu x^\nu}$ that is distinct from the Euclidean norm $|x| = \sqrt{(x^1)^2 + \dots + (x^d)^2}$. We use the subscript b to indicate this second “diffusion” norm. The stationary density satisfies the time-independent Fokker-Planck equation $\mathcal{L}P_1(x) = 0$ and, when it exists, is reached asymptotically in time for arbitrary initial distributions, $\lim_{t \rightarrow \infty} P_{1|1}(x, t|x_0) = P_1(x)$. The stationary distribution is unique for ergodic systems [57].

Associated with the Itô process is the Freidlin-Wentzell “action” functional [2, 3, 4]

$$S[x(t)] = \frac{1}{2} \int_0^T |\dot{x} - a(x)|_b^2 dt \quad (1.3)$$

which gives an asymptotic estimate for the logarithm of the probability of trajectories $X(t)$ to remain in the tubular neighbourhood of a smooth path $x(t)$ over the duration $0 \leq t \leq T$. We write this as

$$P_{\text{tube}}[x(t)] \asymp \exp\left(-\frac{1}{\varepsilon}S[x(t)]\right) \quad (1.4)$$

which, in terms of limits, means

$$S[x(t)] = \lim_{\delta \rightarrow 0} \lim_{\varepsilon \rightarrow 0} -\varepsilon \ln P\left[\sup_{0 \leq t \leq T} |X(t) - x(t)|_b < \delta\right].$$

The limits must be taken in the order above as they do not commute. Eq. 1.4 is a large

deviation principle for trajectories of Itô processes, due to Wentzell and Freidlin and Graham [58].

For reasons described below, it is of interest to obtain the mode of the tube probability over the set of continuous paths

$$\gamma_T = \{x(t) \mid x(0) = x_1, x(T) = x_2, 0 \leq t \leq T\}$$

which have fixed termini x_1 and x_2 and are of duration T . This is equivalent to the variational problem of minimising the Freidlin-Wentzell action. The minimum value of the action,

$$V_T(x_2|x_1) = \min_{\gamma_T} S[x(t)], \quad (1.5)$$

is called the quasipotential. The path attaining the minimum,

$$x_T^*(t) = \arg \min_{\gamma_T} S[x(t)], \quad (1.6)$$

is called the instanton. We emphasise that this path describes the smooth centerline of the tube of maximum probability and not a non-differentiable trajectory of the diffusion process. It is the most probable dynamical path connecting two points in configuration space.

The instanton and the quasipotential are central objects in Freidlin-Wentzell-Graham theory and relate to the eikonal approximation of the Fokker-Planck equation [59]. Assuming the JWKB form of the transition density,

$$P_{1|1}(x, t|x_0) \sim \exp \left[\frac{1}{\varepsilon} \sum_{n=0}^{\infty} \varepsilon^n \phi_n(x, x_0; t) \right],$$

with prefactors suppressed, substituting in the Fokker-Planck equation and matching terms gives a Hamilton-Jacobi equation for the lowest order contribution,

$$\partial_t \phi_0 + \frac{1}{2} b^{\mu\nu} \partial_\mu \partial_\nu \phi_0 + a^\mu \partial_\mu \phi_0 = 0. \quad (1.7)$$

This corresponds to the Hamiltonian system

$$\begin{aligned} H(x, p) &= \frac{1}{2} b^{\mu\nu} p_\mu p_\nu + a^\mu p_\mu \\ \dot{x}^\mu &= + \frac{\partial H}{\partial p_\mu} = b^{\mu\nu} p_\nu + a^\mu \\ \dot{p}_\mu &= - \frac{\partial H}{\partial x^\mu} = - \frac{\partial b^{\nu\lambda}}{\partial x^\mu} p_\nu p_\lambda - \frac{\partial a^\nu}{\partial x^\mu} p_\nu \end{aligned} \quad (1.8)$$

whose solutions define an equivalent variational problem of extremising an action with the Lagrangian

$$\begin{aligned} L(x, \dot{x}) &= p_\mu \dot{x}^\mu - H(x, p) \\ &= \frac{1}{2}(\dot{x}^\mu - a^\mu)b_{\mu\nu}(\dot{x}^\nu - a^\nu). \\ &= \frac{1}{2}|\dot{x} - a(x)|_b^2 \end{aligned} \tag{1.9}$$

Thus, the rays of the Hamilton-Jacobi equation that determine the lowest order contribution to the eikonal are local maxima of the tube probability, or in other words, $\phi_0(x, x_0; T) = V_T(x|x_0)$. The large-deviation principle of Freidlin and Wentzell and the theory of the non-equilibrium potential of Graham [3, 4] thus appear as elegant reformulations of the JWKB approximation [59].

The correspondence with the JWKB approximation yields the asymptotic form of the transition density,

$$P_{1|1}(x, T|x_0) \asymp \exp\left[-\frac{1}{\varepsilon}V_T(x|x_0)\right], \tag{1.10}$$

and, in the $T \rightarrow \infty$ limit of the above, the asymptotic form of the stationary distribution,

$$P_1(x) \asymp \lim_{T \rightarrow \infty} \exp\left[-\frac{1}{\varepsilon}V_T(x|x_0)\right]. \tag{1.11}$$

where x and x_0 belong to the same basin of attraction of an attractor \mathcal{A} . As is indicated by Eq. 1.11, it can be shown that this limit is independent of the initial coordinate,

$$\lim_{T \rightarrow \infty} V_T(x|x_0) = V_\infty^\mathcal{A}(x), \tag{1.12}$$

where $V_\infty^\mathcal{A}$ is equal, to within a constant, to the stationary quasipotential $V_\infty(x)$ in the basin of attraction of \mathcal{A} . For a system with multiple attractors \mathcal{A}_i , the global quasipotential is

$$V_\infty(x) = \min_i (V_\infty^{\mathcal{A}_i}(x) + C^{\mathcal{A}_i}) \tag{1.13}$$

where $C^{\mathcal{A}_i}$ is an additive constant. The constants are fixed by requiring

$$V_\infty^{\mathcal{A}_i}(x_s^{(i,j)}) + C^{\mathcal{A}_i} = V_\infty^{\mathcal{A}_j}(x_s^{(i,j)}) + C^{\mathcal{A}_j} \tag{1.14}$$

for attractors \mathcal{A}_i and \mathcal{A}_j with adjacent basins of attraction, where $x_s^{(i,j)}$ is the saddle with the lowest value on the separatrix between the basins [4]. The stationary quasipotential determines the mean persistence time of a trajectory in a basin of attraction which generalises the Arrhenius law to systems out of equilibrium.

The $T \rightarrow \infty$ limit involved in the definition of the stationary quasipotential presents considerable numerical difficulties in the minimisation of the Freidlin-Wentzell action. A more numerically amenable route to determining the stationary quasipotential is by the minimisation of the action over paths that start at an attractor and end at a point in its basin, regardless of the duration. We show in the next section that the solution of this second variational problem does, indeed, yield the stationary quasipotential and derive an alternative form of the Freidlin-Wentzell action that is adapted to computing instantons regardless of their duration.

1.3 On-shell action

We consider the variational problem of minimising the Freidlin-Wentzell action over paths with fixed termini but of arbitrary duration,

$$\min_T \min_{\gamma_T} S[x(t)] = \min_T \min_{\gamma_T} \int_0^T L(x, \dot{x}) dt, \quad (1.15)$$

where both the initial and final points are in the basin of the attraction \mathcal{A} and the Freidlin-Wentzell Lagrangian following from Eq. 1.9 is

$$L(x, \dot{x}) = \frac{1}{2} b_{\mu\nu} \dot{x}^\mu \dot{x}^\nu - b_{\mu\nu} a^\mu \dot{x}^\nu + \frac{1}{2} b_{\mu\nu} a^\mu a^\nu. \quad (1.16)$$

This variational problem can be solved by introducing a parametrisation u for both the coordinate and time,

$$x = x(u), \quad x' = dx/du; \quad t = t(u), \quad t' = dt/du,$$

that allows the shape of the path,

$$\sigma = \{x(u) \mid x(u_1) = x_1, x(u_2) = x_2, u_1 \leq u \leq u_2\},$$

to be varied independently of its duration,

$$T = \int_{u_1}^{u_2} t' du.$$

Coordinates x and time t are dependent variables in the reparametrised action,

$$S[x(u)] = \int_{u_1}^{u_2} L(x, \frac{x'}{t'}) t' du, \quad (1.17)$$

in which the time-dependence appears only through the derivative t' . Therefore, t is a cyclic (or ignorable) coordinate and Noether's theorem implies that the corresponding conjugate momentum is conserved [60]:

$$-\frac{\partial(Lt')}{\partial t'} = \frac{1}{2}b_{\mu\nu}\frac{x'^\mu x'^\nu}{(t')^2} - \frac{1}{2}b_{\mu\nu}a^\mu a^\nu = E. \quad (1.18)$$

This defines submanifolds of the dynamics labelled by the “energy” E which we shall call shells. The bound $2E + |a|_b^2 \geq 0$ for the energy follows immediately from the positive-definiteness of the diffusion tensor.

Solving the first integral for t' gives

$$t' = \frac{dt}{du} = \frac{|x'|_b}{\sqrt{2E + |a|_b^2}}, \quad (1.19)$$

from which the duration of the path is obtained to be

$$T_E = \int_{u_1}^{u_2} \frac{|x'|_b}{\sqrt{2E + |a|_b^2}} du. \quad (1.20)$$

This shows that paths γ_{T_E} (of duration T_E) are equivalent to shapes σ_E (of energy E), where the latter is the restriction of shapes in σ to the shell of constant energy. Then, minimisation over paths γ_T regardless of their duration is equivalent to minimisation over shapes σ_E regardless of their energy, or

$$\min_T \min_{\gamma_T} S[x(t)] = \min_E \min_{\sigma_E} S[x(u)].$$

The action for shapes restricted to σ_E is obtained by eliminating t' between Eq. 1.17 and Eq. 1.19. This gives the “on-shell” form of the Freidlin-Wentzell action,

$$S_E[x(u)] = \int_{u_1}^{u_2} \left[\frac{E + |a|_b^2}{\sqrt{2E + |a|_b^2}} |x'|_b - b^{\mu\nu} a_\mu x'_\nu \right] du,$$

which is a functional of the shape $x(u)$, a function of the energy E , and allows for independent variations of both. It is straightforward to see that the integrand and, therefore, the action is minimised when $E = 0$. Therefore, most probable paths, regardless of their duration, are obtained by minimising

$$S_0[x(u)] = \int_{u_0}^{u_1} (|a|_b |x'|_b - b^{\mu\nu} a_\mu x'_\nu) du \quad (1.21)$$

over shapes restricted to the zero-energy shell. The duration on the zero-energy shell,

$$T_0 = \int_{u_1}^{u_2} \frac{|x'|_b}{|a|_b} du, \quad (1.22)$$

shows that paths that leave, cross, or terminate at points of vanishing drift, $a^\mu(x) = 0$, are necessarily of infinite duration. The corresponding shapes σ_0^A can then be taken to start at a fixed point and end at another point x in the basin of attraction. The quasipotential is determined by a minimisation over such shapes σ_0^A ,

$$V_\infty^A(x) = \min_{\sigma_0^A} S_0[x(u)], \quad (1.23)$$

and the shape attaining the minimum,

$$x_\infty^*(u) = \arg \min_{\sigma_0^A} S_0[x(u)], \quad (1.24)$$

is the stationary instanton. The time on the instanton path can be obtained by integrating $t' = |x'|_b/|a|_b$. The utility of the on-shell form of the action is that it provides the shape of the path independently of its duration. The latitude of obtaining the shape from a parametrisation over a finite interval, even for paths of infinite duration, is extremely useful in numerical work.

The on-shell action is related to, but distinct from, the Jacobi action in mechanics [61, 62], which, following a Routh reduction [60], would in this case be

$$\hat{S}[x(u)] = \int_{u_1}^{u_2} \left[\sqrt{2E + |a|_b^2} |x'|_b - a^\mu x'_\mu \right] du.$$

Though both the on-shell and Jacobi action agree on the zero-energy shell, only the former supports the interpretation as least action for non-zero energies. Furthermore, variations of the on-shell action have to respect the on-shell condition Eq. 1.19 (in other words, the solutions of its Euler-Lagrange equations does not coincide with its extrema). On the other hand, the Jacobi action can be varied using the standard Euler-Lagrange approach.

For gradient dynamics, that is $a^\mu = b^{\mu\nu} \partial U / \partial x^\nu$, the on-shell action generalises the “scalar work” functional of Olender and Elber [63] to non-zero energies and configuration-dependent diffusion tensors. For non-gradient dynamics, where the drift cannot be so expressed, the on-shell action generalises the geometric action of Heyman and Vanden-Eijnden [64] to non-zero energies. The non-zero energy shell $|\dot{x}|_b^2 = |a|_b^2 + E$ admits the most general path consistent with time-translation invariance, in contrast to the zero-energy shell where the magnitude of the velocity is always equal to that of the drift, $|\dot{x}|_b^2 = |a|_b^2$. Such general paths determine the quasipotential and the asymptotic form of the transition density for finite times and will be examined in detail in future work. Accordingly, we set $E = 0$ below. The derivation of the on-shell action only requires time-translation invariance of the Lagrangian and not, as in [63, 64], their positive-definiteness. Thus, it can be applied to stochastic actions whose Lagrangians are not necessarily positive-definite, as for example the Onsager-Machlup action [25, 29]. We now describe the Ritz method by

which we minimise actions.

1.4 Ritz method

The direct method in the calculus of variations consists of constructing a sequence of extremisation problems for a function of a finite number of variables that, in the passage to the limit of an infinite number of variables, yields the solution to the variational problem. The two main families of direct methods are finite differences and Ritz methods [46, 47, 49, 65]. In the latter, the solution of the variational problem is sought in a sequence of functions

$$\varphi_1(t), \varphi_2(t), \dots, \varphi_n(t), \dots$$

each of which satisfies end point conditions. The path is expressed as a linear combination of these functions

$$x_n(t) = \alpha_1 \varphi_1(t) + \dots + \alpha_n \varphi_n(t) \quad (1.25)$$

which transforms the action from a functional of the path into a function of the expansion coefficients,

$$\begin{aligned} S(\alpha_1, \dots, \alpha_n) &= \int_0^T L(x_n, \dot{x}_n) dt \\ &= \int_0^T L\left(\sum_{i=1}^n \alpha_i \varphi_i, \sum_{i=1}^n \alpha_i \dot{\varphi}_i\right) dt. \end{aligned} \quad (1.26)$$

Action minimisation now becomes a search for a set of coefficients, α_i^* such that $S(\alpha_1^*, \dots, \alpha_n^*) < S(\alpha_1, \dots, \alpha_n)$. The necessary condition for this is the vanishing of the gradient,

$$\frac{\partial S}{\partial \alpha_i} = 0 \quad (i = 1, 2, \dots, n), \quad (1.27)$$

which is the Ritz system of non-linear equations. Coefficients satisfying these conditions can be obtained by non-linear optimisation. The n -th approximation to the minimum action path, $x_T^*(t)$, and the minimum value of the action, $S[x_T^*(t)]$, are obtained from these values of the coefficients. It is generally the case that this sequence of approximations converges to the minimum of the variational problem as $n \rightarrow \infty$ [46, 47].

The method, then, has three parts: first, the choice of basis functions $\varphi_i(t)$; second, the quadrature rule that integrates the Lagrangian to obtain the action as a function of the expansion coefficients; and third, the optimisation that yields the coefficients at the minimum. Since each part is only loosely dependent on the others, Ritz methods come in many varieties [66]. Our choices are centered around Chebyshev polynomials as described below. Approximation by Chebyshev polynomials and their optimality for the purpose are

described in [67, 68, 69].

Basis functions: We consider a path $x(u)$ that is a Lipschitz continuous function of the parameter u in the interval $[-1, 1]$. Then, it has an absolutely and uniformly convergent Chebyshev expansion,

$$x(u) = \sum_{k=0}^{\infty} a_k T_k(u), \quad a_k = \frac{2}{\pi} \int_{-1}^1 \frac{x(u) T_k(u)}{\sqrt{1-u^2}} du$$

where $T_k(u)$ are Chebyshev polynomials of the first kind and the integral must be halved for $k = 0$. A suitable sequence of paths can be constructed from the first n terms of this infinite series. However, it is computationally more convenient, for reasons that will be clear below, to construct the sequence from n -th degree polynomials that interpolate the path at the $n + 1$ Chebyshev points

$$u_j = -\cos(j\pi/n), \quad (j = 0, 1, \dots, n). \quad (1.28)$$

The n -th degree interpolant can be expressed in standard form as a sum of Lagrange cardinal polynomials $\ell_j(u)$ or as a linear combination of Chebyshev polynomials,

$$x_n(u) = \sum_{j=0}^n \alpha_j \ell_j(u) = \sum_{k=0}^n c_k T_k(u). \quad (1.29)$$

The coefficients c_k are aliased versions of the coefficients a_k . Since the cardinal polynomials have the property

$$\ell_j(u_k) = \begin{cases} 1, & j = k \\ 0, & \text{otherwise,} \end{cases} \quad (j, k = 0, \dots, n),$$

$x_n(u_k) = \alpha_k$, that is, the expansion coefficients α_k are path coordinates at the Chebyshev points. Expressing the entire path in terms of its discrete coordinates has the advantage that end point conditions can be imposed by setting

$$\alpha_0 = x(u_0) = x_0, \quad \alpha_n = x(u_n) = x_1. \quad (1.30)$$

Admissible paths of degree n are, then, parametrised by the $n - 1$ independent coefficients $\alpha_1, \dots, \alpha_{n-1}$. In contrast, imposing end point conditions in series form leads to a numerically inconvenient linear dependence between the coefficients c_k . The derivative of the path is a polynomial of degree $n - 1$ that can be expressed in terms of the interpolant as

$$x'_n(u) = \sum_{j=0}^n \alpha_j \ell'_j(u) = \sum_{j=0}^n \beta_j \ell_j(u) \quad (1.31)$$

with the two sets of expansion coefficients related by the Chebyshev spectral differentiation matrix

$$\beta_j = D_{jk}\alpha_k, \quad D_{jk} = \ell'_k(u_j). \quad (1.32)$$

We use the barycentric form of the Lagrange polynomials [70]

$$\ell_j(u) = \frac{w_j}{u - u_j} \Big/ \sum_{k=0}^n \frac{w_k}{u - u_k}. \quad (1.33)$$

with weights [71]

$$w_j = \begin{cases} \frac{1}{2}, & j = 0 \\ (-1)^j, & j = 1, \dots, n-1 \\ \frac{1}{2} \cdot (-1)^n, & j = n. \end{cases}$$

This form is both numerically stable and, costing no more than $O(n)$ operations, efficient to evaluate. [72].

Chebyshev interpolants converge exponentially for analytic functions and algebraically for functions with a finite number of derivatives. More precisely, for an analytic path, $\|x - x_n\| = O(\rho^{-n})$ for some $\rho > 1$ as $n \rightarrow \infty$. For a path with ν derivatives and ν -th derivative of bounded variation K , $\|x - x_n\| = O(Kn^{-\nu})$ as $n \rightarrow \infty$. These estimates are in the supremum norm $\|a\|$, that is, the maximum of the absolute value of a in the interval $[-1, 1]$. In contrast, finite-difference methods can only achieve polynomial, but never exponential, rates of convergence, even for analytic paths [67, 68].

Quadrature: To reduce the action to a multivariate function of the coefficients it is necessary to evaluate the integral

$$S(\alpha_1, \dots, \alpha_n) = \int_{-1}^1 L(x_n(u), x'_n(u)) du \quad (1.34)$$

using a quadrature rule. For instance, quadrature at the Chebyshev points u_j gives

$$\begin{aligned} S(\alpha_1, \dots, \alpha_n) &= \sum_{j=0}^n \omega_j L(x_n(u_j), x'_n(u_j)) \\ &= \sum_{j=0}^n \omega_j L(\alpha_j, \beta_j) \\ &= \sum_{j=0}^n \omega_j L(\alpha_j, D_{jk}\alpha_k) \end{aligned}$$

where ω_j are the quadrature weights. However, standard quadrature rules at this set of n Chebyshev points, which integrate a polynomial of degree less than or equal to n exactly, will generally be inaccurate. The reason is that the Lagrangian has polynomial degree

different from, and usually greater than, the polynomial degree of the path. For instance, when b_{ij} is a constant, the term quadratic in the velocities has twice the polynomial degree of the path. Therefore, if the Lagrangian is to be integrated accurately, the order of the quadrature must be different from, and in general greater than, the polynomial degree of the path.

Therefore, we define a second set of $n_q > n$ Chebyshev points

$$v_j = -\cos(j\pi/n_q), \quad (j = 0, 1, \dots, n_q) \quad (1.35)$$

and interpolate the path at these points. This is done efficiently by matrix multiplication with a $(n_q + 1) \times (n + 1)$ matrix

$$x_n(v_j) = \sum_{k=0}^{n_q} B_{jk} \alpha_k, \quad (1.36)$$

$$x'_n(v_j) = \sum_{k=0}^{n_q} B_{jk} \beta_k, \quad (1.37)$$

whose elements are derived from the barycentric interpolant

$$B_{jk} = \frac{w_k}{v_j - u_k} \Big/ \sum_{l=0}^{n_q} \frac{w_l}{v_l - u_l}. \quad (1.38)$$

The Lagrangian is evaluated at these second set of points after which Clenshaw-Curtis quadrature [67, 68] is used to evaluate the action,

$$\begin{aligned} S(\alpha_1, \dots, \alpha_n) &= \sum_{j=0}^{n_q} \omega_j L(x_n(v_j), x'_n(v_j)) \\ &= \sum_{j=0}^{n_q} \omega_j L(B_{jk} \alpha_k, B_{jk} \beta_k) \\ &= \sum_{j=0}^{n_q} \omega_j L(B_{jk} \alpha_k, B_{jk} D_{kl} \alpha_l) \\ &\equiv \sum_{j=0}^{n_q} \omega_j L(B_{jk} \alpha_k, C_{jk} \alpha_k). \end{aligned} \quad (1.39)$$

As with interpolation, Clenshaw-Curtis quadrature converges exponentially for Lagrangians that are analytic in u and algebraically for Lagrangians with a finite number u derivatives. Precise estimates are given in [69]. For fixed values of n and n_q , the matrices B_{ij} and C_{ij} in the above expression are constant and can be precomputed and stored. The multiplications

require $O(nn_q)$ operations, and so there is a linear cost, for fixed n , to increase the order of the quadrature. For Lagrangians of polynomial order n_L , the number of quadrature points must be $n_q > (n + 1)n_L$. For nonpolynomial Lagrangians, n_q has to be chosen to ensure that the n_q -th Chebyshev coefficient is suitably small. Well-defined procedures exist for the adaptive truncation of Chebyshev series [73] but here we use a simple rule of thumb and set $n_q = 10n$ leaving the implementation of more efficient truncations to future work. We note that in the direct finite-difference method, introduced in [48], the path is interpolated at uniformly spaced points by a quadratic polynomial and the Lagrangian is integrated using the trapezoidal rule. This combination can exactly evaluate the action for Lagrangians that are at most quadratic polynomials.

Optimisation: To minimise the action over the expansion coefficients $\alpha_1, \dots, \alpha_{n-1}$ we use both gradient-free and gradient-based algorithms. For gradient-free algorithms we provide Eq. 1.39 directly. For algorithms that require the gradient, the chain rule gives

$$\begin{aligned}\nabla_{\alpha_i} S &= \frac{\partial}{\partial \alpha_i} \left[\sum_{j=1}^{n_q} \omega_j L(B_{jk}\alpha_k, C_{jk}\alpha_k) \right] \\ &= \sum_{j=1}^{n_q} \left[\frac{\partial L}{\partial x_n(v_j)} B_{ji}^* + \frac{\partial L}{\partial x'_n(v_j)} C_{ji}^* \right]\end{aligned}\tag{1.40}$$

where $B_{ij}^* = \omega_i B_{ij}$ and $C_{ij}^* = \omega_i C_{ij}$. These matrices, too, can be precomputed and stored and only the partial derivatives of the Lagrangian need to be computed for given values of the coefficients. For the examples presented below, we use *NEWUOA* [74] for gradient-free optimisation and *SLSQP* algorithm [75] for gradient-based optimisation, both of which are implemented in the *NLOPT* numerical optimisation package [76]. For non-equilibrium systems, instantons lose smoothness when passing through fixed points. For such paths, convergence is still achieved but at less than spectral rates. Spectral convergence can be recovered if paths are evaluated piecewise, taking care to isolate the points of derivative discontinuities. This is feasible because fixed points are the only locations where Freidlin-Wentzell instantons can lose smoothness [4].

1.5 Numerical results

In this section, we apply the Ritz method to three diffusion processes that are widely used to benchmark rare event algorithms. The first is overdamped Brownian motion in a complex energy landscape, the second is overdamped Brownian motion under the influence of a circulatory force, and the third is a model of the weather. All three models have configuration-independent diffusion tensors for which it is not necessary to distinguish between covariant and contravariant indices. Python codes for each of these examples are

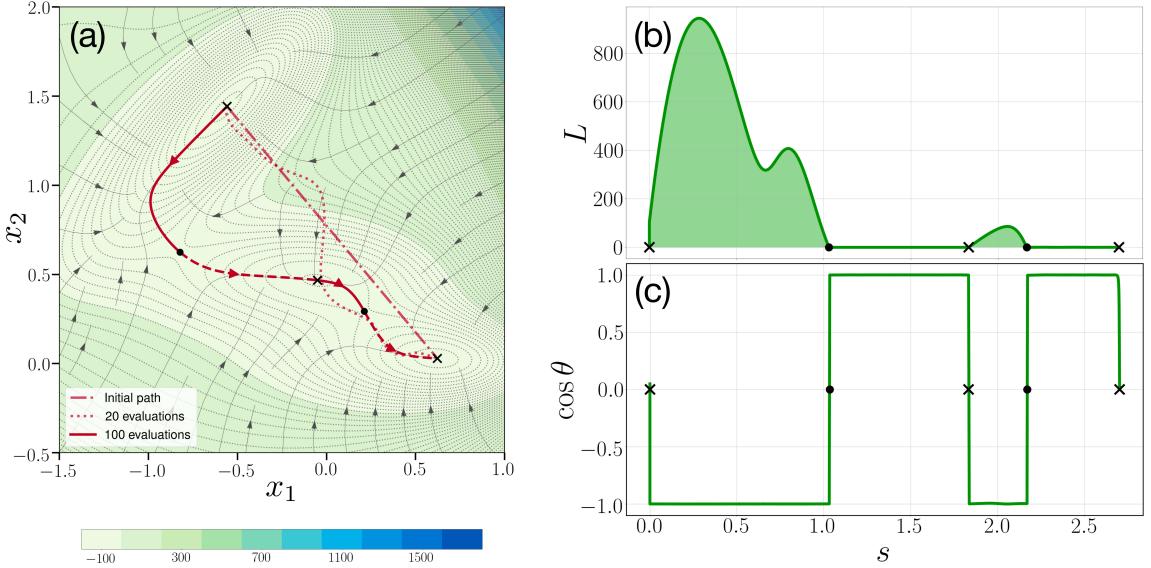


Figure 1.1: Ritz method for overdamped motion in the Müller-Brown potential, which has three minima (crosses) and two saddle points (dots). The initial path is the straight line connecting two minima and the instanton is the solid line, with broken segments showing motion along the force. The instanton automatically locates and passes through both saddles. A typical path before convergence to the minimum is shown as a dotted line. (b) The value of the Lagrangian as a function of Euclidean arc-length of the instanton. The action vanishes to machine precision on segments of the path where motion is along the force. (c) The cosine of the angle θ between the tangent and the force is always ± 1 , i.e. the instanton is a minimum energy path. The instanton is represented by a polynomial of degree $n = 10$.

freely available on GitHub.

1.5.1 Brownian dynamics in a complex potential

Our first example considers the overdamped Brownian motion in a two-dimensional potential with a constant friction. The usual equations of Brownian dynamics can be recast into Itô form,

$$\begin{aligned} dX_1 &= -\mu \partial_1 U dt + \sqrt{2\mu\varepsilon} dW_1 \\ dX_2 &= -\mu \partial_2 U dt + \sqrt{2\mu\varepsilon} dW_2, \end{aligned}$$

where μ is the mobility and $\varepsilon = k_B T$ is the temperature. The Freidlin-Wentzell action for a smooth path with two-dimensional coordinate $x = (x_1, x_2)$ is

$$S[x] = \frac{1}{2} \int_0^T \frac{1}{2\mu} |\dot{x} + \mu \nabla U|^2 dt$$

where $\nabla U = (\partial_1 U, \partial_2 U)$. The minimum of the zero-energy action,

$$S_0[x] = \int_{-1}^1 |\nabla U(x)| |x'| du + [U(x)]_{-1}^1,$$

provides the most probable shape and the stationary quasipotential. The second term does not affect the minimisation and can be discarded. The resulting reduced action

$$\tilde{S}[x] = \int_{-1}^1 |\nabla U| |x'| du \quad (1.41)$$

is of the same form as Fermat's principle for optical rays, where $|\nabla U(x)|$ plays the role of the refractive index and $|x'|du = ds$ is the arc-length of the ray. In geometric optics, Fermat's principle is equivalent to Huygen's principle and its “wavelet equation”

$$\partial_i U = |\nabla U| \frac{dx_i}{ds}. \quad (1.42)$$

This can be easily verified by differentiating it with respect to arc-length, to obtain the eikonal equation

$$\partial_i |\nabla U| = \frac{d}{ds} \left[|\nabla U| \frac{dx_i}{ds} \right],$$

which is identical to the Euler-Lagrange equation of the zero-energy action. The wavelet equation implies that the tangent $t = dx/ds$ to the path is parallel to the gradient of the potential, or equivalently, that rays are normal to contours of the potential. This is the well-known condition for a minimum energy path and was first derived variationally from the scalar work functional by Olender and Elber [63]. It provides a stringent test of the fidelity of the paths obtained by minimisation.

Following [63], we choose the Müller-Brown potential of [77] as an example of a complex energy landscape. The potential and its stationary points are shown in Fig. 1.1. The three minima are marked by crosses and two saddle points by dots. The instanton is computed by requiring the path to start at the minimum on the top left and terminate at the minimum on the bottom right. The initial straight line shape, an intermediate shape and the converged instanton are shown in panel (a). The minimisation automatically locates the two saddle points and makes the instanton pass through them. The action cost along the path is shown in panel (b), where the vanishing of the action on segments of the path along the force is clearly seen. The cosine of the angle between the tangent and force is shown in panel (c) and the condition for a minimum energy path is clearly fulfilled. We emphasise that the condition is not imposed separately but is satisfied automatically at the minimum. The Ritz method provides an alternative to chain-of-states methods for finding minimum energy paths. It does not need the Hessian of the potential, which makes it suitable for problems where such evaluations are expensive. Unlike [64], our parametrisation has no

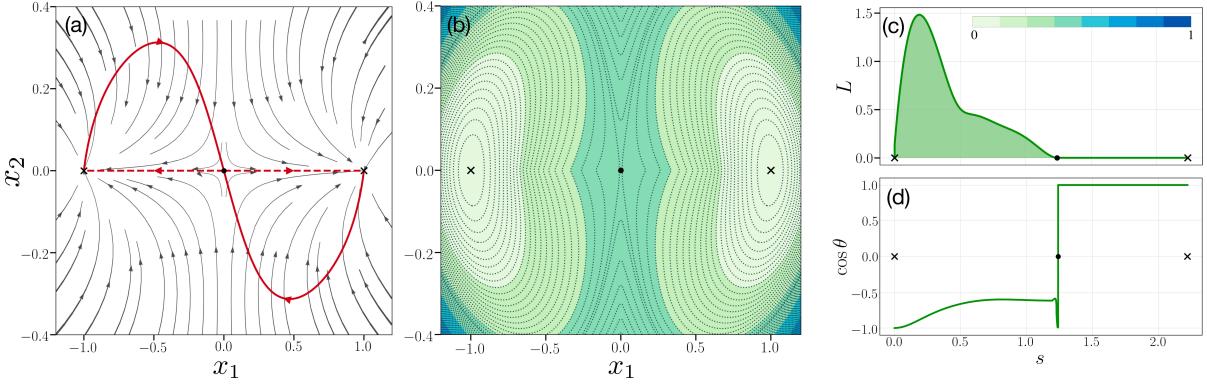


Figure 1.2: Ritz method for overdamped motion in a circulatory (i.e. non-gradient) force field. The instanton is in red with solid (dashed) segments showing motion against (along) the force field. The instanton is reflected about the horizontal axis for motion starting on the right, showing the inequivalence of fluctuational and relaxational paths for non-gradient dynamics. (b) The quasipotential, computed using Eq. 1.13, with a caustic at the unstable fixed point. (c) The value of the Lagrangian as a function of the Euclidean arc-length of the instanton. As in the potential case, the action vanishes to machine precision on segments where motion is along the force. (d) The cosine of the angle θ between the tangent and the force is, unlike in the potential case, not always ± 1 . The instanton is represented by a polynomial of degree $n = 8$.

unit-speed constraint and the minimisation, accordingly, is unconstrained. The method applies without change to dynamics with configuration-dependent friction.

1.5.2 Brownian dynamics in a circulatory field

For our second example we consider, in contrast to the first, Brownian motion in a force field that cannot be derived from a potential and, as such, necessarily has a non-vanishing curl. Choosing the force field of Maier and Stein [12] gives

$$\begin{aligned} dX_1 &= (X_1 - X_1^3 - \beta X_1 X_2^2) dt + \sqrt{\epsilon} dW_1 \\ dX_2 &= -(1 + X_1^2) X_2 dt + \sqrt{\epsilon} dW_2 \end{aligned}$$

for the overdamped motion of the two-dimensional coordinate $X = (X_1, X_2)$, where β is a parameter. The force field $f(x_1, x_2) = (x_1 - x_1^3 - \beta x_1 x_2^2, -(1 + x_1^2)x_2)$ is smooth, and f_1 is odd in x_1 and even in x_2 , while for f_2 the converse holds. There are two stable fixed points at $x_a = (-1, 0)$ and $x_b = (1, 0)$, and a saddle point at $x_s = (0, 0)$. The force field admits a potential only for $\beta = 1$, when it can be written as $f = -\nabla U$, with $U(x_1, x_2) = -\frac{1}{2}x_1^2 + \frac{1}{4}x_1^4 + \frac{1}{2}(1 + x_1^2)x_2^2$. The force field is shown in the first panel of Fig. 1.2 for $\beta = 10$ together with the instanton moving from x_a to x_b . As before, solid (dashed) segments represent motion against (along) the vector field. The instanton moving from x_b to x_a is obtained by reflection about the x_1 -axis showing that that fluctuational and relaxational paths are not identical in a non-gradient field.

The middle panels shows the stationary quasipotential $V_\infty^{\mathcal{A}_i}(x)$ with respect to the attractors at $(-1, 0)$ and $(1, 0)$ respectively. The quasipotential is sampled on a 128×128 grid by computing instantons between a point on the grid and the relevant attractor. The contours of the quasipotential and its heatmap are obtained from these discrete samples. To the best of our knowledge, all prior estimations of the quasipotential for this problem (and more generally, for circulatory forces) have required numerical solutions of the Hamilton-Jacobi equation. Our method of direct sampling provides an alternative to this route of computing the quasipotential. The right panel shows the Lagrangian as a function of arc-length along the instanton. As in the previous example, the Lagrangian vanishes along segments of the path where motion is along the force. For motion against the force, the tangent to the path is no longer parallel to the force, as shown by the variation of the cosine of the angle θ between the tangent and the force. We note that our method is agnostic to the existence, or not, of a potential for the drift and treats both these cases on equal footing.

1.5.3 Multistability in a genetic switch system

We continue with the dynamics of a two-dimensional coordinate $x = (x^1, x^2)$ in a non-gradient field, but now of non-mechanical origin and non-polynomial form,

$$\begin{aligned} dx^1 &= \left(\frac{a_1}{1 + \left(\frac{x^2}{K_2}\right)^n} - \frac{x^1}{\tau} \right) dt + \sqrt{\epsilon} dW^1 \\ dx^2 &= \left(\frac{a_2}{1 + \left(\frac{x^1}{K_1}\right)^m} - \frac{x^2}{\tau} \right) dt + \sqrt{\epsilon} dW^2, \end{aligned} \tag{1.43}$$

where $a_1, a_2, \tau, n, m, K_1$ and K_2 are constants. This model is due to Roma *et al* [78] and describes a multistable genetic network. In the region of configuration space we consider, the vector field has two fixed points one of which is stable and the other a saddle. The top panel of Fig. 1.3 shows the instantons moving between these fixed points. Unlike in the previous examples, here the instantons move either entirely with (solid line) or entirely against (dashed line) the flow. The bottom panel shows the quasipotential with respect to the stable fixed point. The procedure to evaluate and plot it is as described above. The quasipotential provides a quantification of the dispersion of the coordinate about the stable fixed point and a measure of the non-equilibrium ‘‘temperature’’ on the state-space of this non-mechanical system.

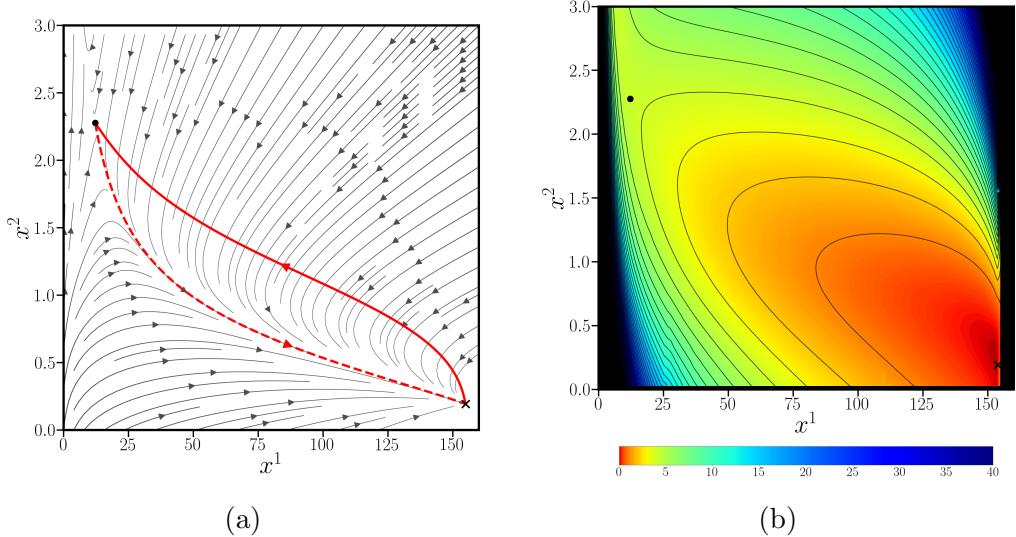


Figure 1.3: Instantons for the genetic switch. (a) shows the instantons in red with solid (dashed) lines representing motion against (along) the vector field. (b) shows the quasipotential with respect to the stable fixed point (cross). Parameter values are $a_1 = 156$, $a_2 = 30$, $\tau = 1$, $n = 3$, $m = 1$ and $K_1 = K_2 = 1$. A polynomial of degree $n = 10$ was used to parametrise the path.

1.5.4 Transitions between limit cycles

To demonstrate that our method is not limited to fixed points, we constructed a simple but non-trivial system:

$$\begin{aligned} dr &= (1 + \cos^2 \theta)f(r)dt + dW^r \\ d\theta &= r(1 + \sin^2 \theta)dt + dW^\theta \end{aligned} \tag{1.44}$$

where $f(r) = -\frac{1}{4}r(r - s_1)(r - s_2)(r - s_3)$, and $s^{(i)} = [1, 3, 5]$. We have an unstable fixed point at $r = 0$, stable limit cycles at $r = s_1$ and $r = s_3$, and an unstable limit cycle at $r = s_2$. The trigonometric factors in the drift breaks the circular symmetry of the system, but preserves the concentric circular limit cycles. We consider instantons moving from the inner stable limit cycle $r = s_1$ to the outer limit cycle $r = s_3$. Let $\Gamma_i = \{(r, \theta) \mid r = s_i\}$ for $i = 1, 2, 3$, be the set of points comprising the three limit cycles. For a given path $x(t)$, let $x_i \in \Gamma_i$ be the points along the path located along the respective limit cycles. Since $a_r(x_i) = 0$ and $a_\theta(x)$ is positive definite, the system can move to any point within a limit cycle without incurring any action cost. Therefore the starting, intermediate and end points x_1 , x_2 and x_3 should be varied freely within their respective limit cycles during the minimisation. We can split the instanton $x^*(t)$ into an ‘‘uphill’’ path $x_{\uparrow}^*(t)$, moving between Γ_1 and Γ_2 , and a ‘‘downhill’’ path $x_{\downarrow}^*(t)$, moving between Γ_2 and Γ_3 . The downhill path follows deterministic relaxational dynamics, and does not contribute to the action, and therefore the only non-trivial part of the problem is the uphill path $x_{\uparrow}^*(t)$. One issue with limit cycle problems is that instantons in general have infinite arc-lengths. In the

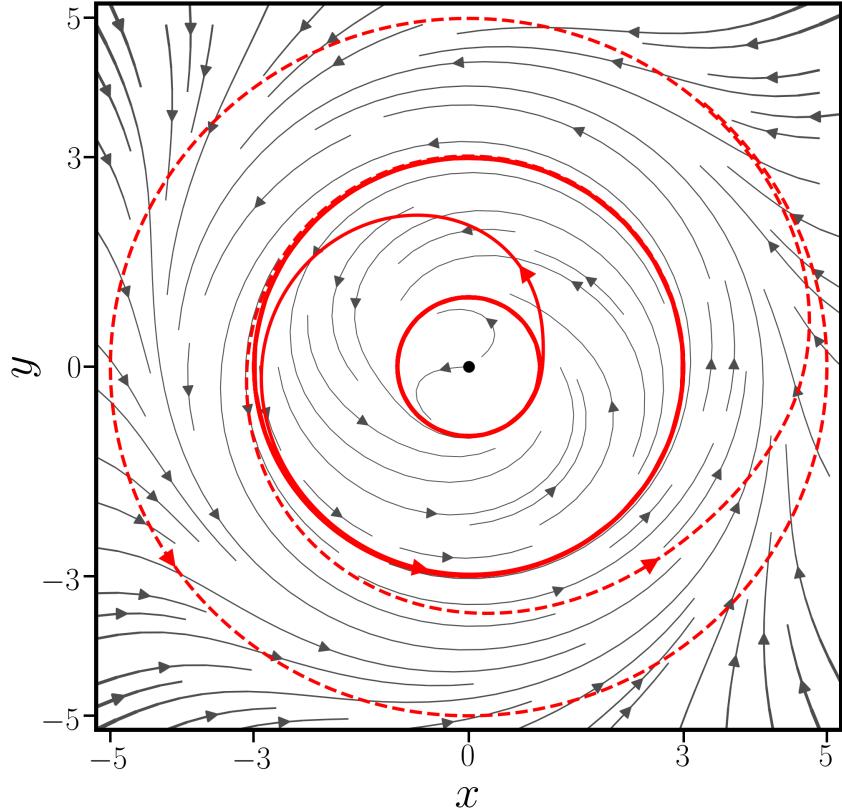


Figure 1.4: An approximate instanton of the concentric limit cycle system with $S \approx 3.65053$. The instanton moves from the inner stable limit cycle at $r = 1$ to the unstable limit cycle at $r = 3$, and then moves hetero-clinically along the drift to the outer stable limit cycle at $r = 4$.

case of the relaxational path $x_{\downarrow}^*(t)$, this can be verified to be the case using an ODE solver. The system will not only relax to the attractor in infinite time, but the system will also undergo an infinite number of cycles before reaching the stable limit cycle Γ_3 . We would expect similar behaviour for diffusive paths leaving stable limit cycles. Paths of infinite Euclidean arc-length are not possible to parametrise exactly in the Chebyshev basis, so only an approximate finite-length instanton can be found, as shown in Fig. 1.4. As we increased the arc-length the action of the candidate instanton converged to a value of $S \approx 3.65053$.

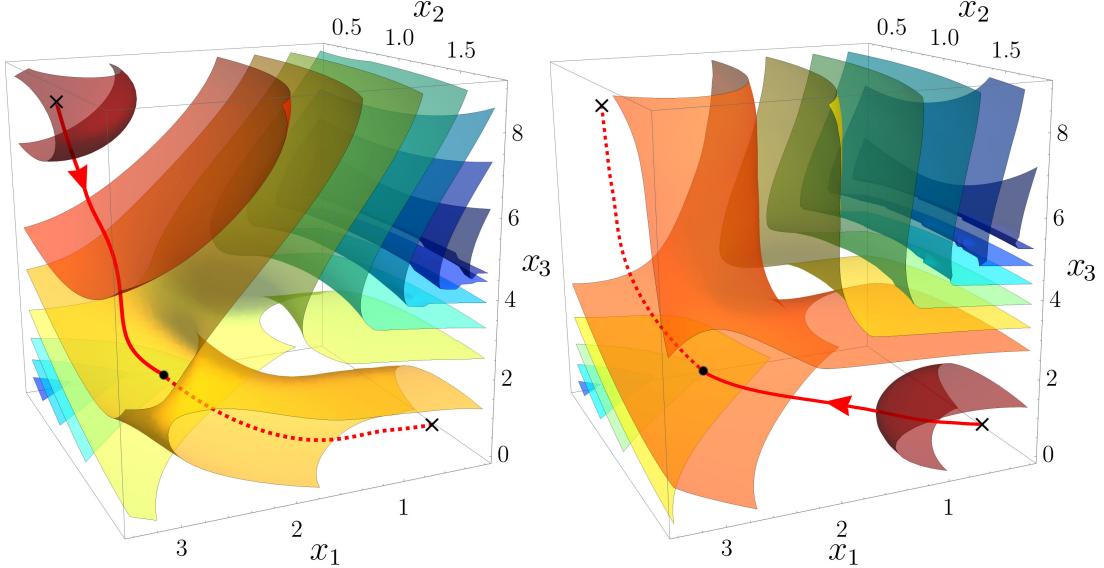


Figure 1.5: Instantons and quasi-potentials of the Egger model. The instantons are shown in red with solid (dashed) lines representing motion against (along) the vector field. The left and right panels are forward and reverse instantons. Isosurfaces of the quasipotential with respect to each attractor is shown in the respective panels. Isovalues increase from light red to blue in the range $\{1, 7, 11, 16, 21, 26, 31, 36\}$. Parameter values are $k = 2$, $\beta = 1.25$, $\gamma = 2$, $U_0 = 10.5$ and $H = 12$. The instanton is represented by a polynomial of degree $n = 10$.

1.5.5 Egger model of weather

Our final example is a reduced model of the weather for a three-dimensional coordinate $X = (X_1, X_2, X_3)$ that has a circulatory drift,

$$\begin{aligned} dX_1 &= \left[kX_2 \left(X_3 - \frac{\beta}{k^2} \right) - \gamma X_1 \right] dt + \sqrt{\epsilon} dW_1 \\ dX_2 &= \left[kX_1 \left(\frac{\beta}{k^2} - X_3 \right) - \gamma X_2 + \frac{HX_3}{k} \right] dt + \sqrt{\epsilon} dW_2 \\ dX_3 &= \left[-\frac{1}{2}HkX_2 - \gamma(X_3 - U_0) \right] dt + \sqrt{\epsilon} dW_3 \end{aligned} \quad (1.45)$$

where k , β , γ , U_0 and H are constants. This model is due to Egger [79]. It is not particularly illuminating to visualise the three-dimensional vector field describing this dynamics but we note that it has two stable fixed points, marked by crosses in Fig. 1.5, and a saddle fixed point marked by a dot. The instanton moving between these points is shown as before in the left and right panels of the figure. Also shown are isosurfaces of the quasipotential with respect to the stable fixed points, with isovalues increasing from red to blue. To the best of our knowledge, this is the first computation of the quasipotential for this model. We provide this example primarily to demonstrate the feasibility of sampling quasipotentials in dimensions greater than two with our method.

Model	$S_1 - S_2$	$S_3 - S_4$	$S_7 - S_8$	$S_{15} - S_{16}$	$S_{31} - S_{32}$
M-B	2	3×10^{-4}	1×10^{-7}	5×10^{-14}	1×10^{-13}
M-S	2×10^{-3}	3×10^{-7}	2×10^{-12}	1×10^{-16}	5×10^{-16}
Egger	8×10^{-3}	1×10^{-3}	6×10^{-7}	4×10^{-9}	8×10^{-13}

Model	$S_2 - S_{50}$	$S_4 - S_{50}$	$S_8 - S_{50}$	$S_{16} - S_{50}$	$S_{32} - S_{50}$
M-B	4×10^{-2}	4×10^{-5}	8×10^{-8}	1×10^{-12}	3×10^{-13}
M-S	2×10^{-4}	7×10^{-11}	2×10^{-13}	1×10^{-16}	2×10^{-16}
Egger	5×10^{-3}	1×10^{-4}	1×10^{-6}	1×10^{-8}	2×10^{-10}

Table 1.1: Convergence of the action S_n for a path of polynomial order n . The abbreviations M-B and M-S refer to Brownian dynamics in the Müller-Brown potential and the Maier-Stein force field respectively. The first table shows the difference $S_n - S_{n+1}$ while the second table shows the difference $S_n - S_{50}$. A tenth-order polynomial typically gives at least six digits of accuracy.

1.6 Numerical convergence

We briefly recall the convergence properties of the Ritz method, comprising that of the basis functions, the quadrature, and the optimisation. The Chebyshev interpolant converges to the most probable path, assuming that it is Lipschitz continuous, at a rate that increases with the number of derivatives the path admits and is exponential for a smooth path. Likewise, the Clenshaw-Curtis quadrature is guaranteed to converge to the minimum of the action, assuming that the Lagrangian is Lipschitz continuous. The optimal number of quadrature points for accuracy to machine precision can be obtained by following the decay of the Chebyshev coefficients of the Lagrangian and truncating at that value beyond which the coefficients vanish to machine precision. The optimisation has lesser theoretical guarantees than the interpolation and quadrature, as is generally the case with search in high-dimensional spaces. However, the residual of the Ritz system provides an empirical measure for how closely the minimum has been located. In all three examples (and in others not presented here) we have found both gradient-free and gradient-based optimization to robustly locate the minima, and gradient-based methods to yield faster convergence. We note that for equilibrium problems, the gradient-free method does not require the Hessian of the energy function, which can be of significant computational advantage. In Table 1.1 we show the spectral convergence of the action with increasing polynomial order of the path for each of our examples.

1.7 Conclusion

We have presented an efficient and accurate numerical method for computing most probable transitions paths and quasipotentials of rare diffusive events. The method directly minimises the Freidlin-Wentzell action and thus provides a unified approach for transition

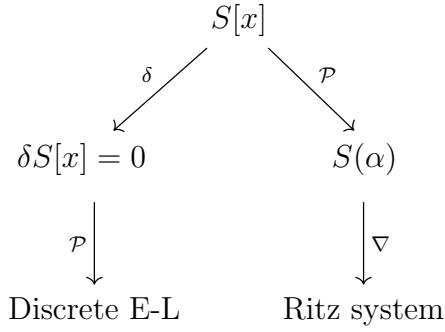


Figure 1.6: Inequivalence of the direct and Euler-Lagrange routes to numerical action minimization. Here, $\delta \rightarrow$ functional variation, $\mathcal{P} \rightarrow$ finite-dimensional projection, and $\nabla \rightarrow$ function minimisation. On the left branch, the action is first varied to obtain the Euler-Lagrange equation and then projected onto a finite-dimensional basis for numerical solution. On the right branch, the action is first projected onto a finite-dimensional basis and then minimised to obtain the Ritz system. The finite-dimensional projection of the Euler-Lagrange equations is, in general, not identical to the Ritz system.

paths in both equilibrium and non-equilibrium systems in the zero-temperature limit. Our reparametrisation-invariant form of the action, derived using a Noether symmetry, is well-suited for numerical work and is a generalisation of the geometric action. This frees us from the constraints of the commonly used arc-length path parametrisation and offers the maximum flexibility in choosing the space of polynomials in which action is minimised. Thus our method is not limited to the Chebyshev polynomials in $[-1, 1]$ used here but easily admits trigonometric polynomials and, more generally, any global basis. Numerical quadrature reduces the action to a multivariate function of coefficients of the path polynomial whose minimum is obtained by both gradient-free and gradient-based optimisation. This gives, simultaneously, both the minimum value of the action and the most probable path. This efficiency of the method allows us to repeatedly compute minimum action paths between an attractor and a point in its basin of attraction and, thereby, map out the quasipotential. The quasipotential in a non-equilibrium steady state has the same significance as the Gibbs distribution in equilibrium and our method provides a robust way of obtaining it without the need to numerically solve the Hamilton-Jacobi partial differential equation.

The direct method used here consists of a discretisation of the action followed by a search for the minimum in the resulting finite-dimensional space, expressed schematically in Fig. (1.6). In contrast, the majority of methods impose the vanishing variation of the action and then search for the solution of the Euler-Lagrange equation in a finite-dimensional space. The resulting discretised Euler-Lagrange equations is, in general, not identical to the Ritz system; in other words, these two methods of reducing an infinite-dimensional problem to a finite-dimensional one are not equivalent. In contrast to mechanics, where

Newton's equations of motion are considered primary and the action derived, here it is the tube probability and hence the Freidlin-Wentzell action that is primary and the Euler-Lagrange equation for the most probable path that is derived. It appears more natural to us to discretise the primary, rather than the derived, object directly. Our approach is algorithmically simple and the only adjustable parameters are the polynomial order n and the quadrature order n_q . This simplicity does not compromise accuracy or efficiency, as confirmed by our examples.

In subsequent chapters we will move beyond the Freidlin-Wentzell regime, which corresponds to the limit of zero diffusivity and infinite duration, to study transition paths in systems at finite temperatures and of varying durations. In this setting, the Ritz methods developed here can be applied to find finite-temperature instantons of the *Onsager-Machlup action*. Although in this regime the action does not admit a on-shell form, as was the case for the Freidlin-Wentzell action.

Chapter 2

Monte Carlo methods on path spaces

2.1 Introduction

In a large deviation limit, the dominant transition pathway of an Itô diffusion equation between fixed points of its drift field is given by the minimiser of the associated Freidlin-Wentzell action, which is known as the instanton of the system. Physically, we found the instanton in a limit of vanishing temperature. Although useful for many applications (as discussed in the introduction of the previous chapter), in any real system the zero-temperature limit is un-physical. There may be multiple locally most-probable paths - *local instantons* - and at finite temperatures the fluctuations around these become increasingly relevant [33, 34, 35, 36]. Furthermore, in some noise-regimes the sample realisations of the system will follow stochastic trajectories that do not necessarily concentrate around nor coincide with the instantons. There may be multiple competing transition channels (as will be a major topic in Ch. 3), or no clearly defined transition channels whatsoever. Therefore at finite temperatures, it is of relevance to consider the *transition path ensemble* (TPE), the set of all transition paths of a system given fixed end-points.

Due to the fixed end-points, it is often inefficient to use stochastic ODE solvers (like the Euler-Maruyama integrator [80]) to sample the transition path ensemble. Furthermore, in low-diffusivity regimes, sampling using such methods is intractable. In this chapter we will apply some recent mathematical developments in the field of infinite-dimensional *Markov-Chain Monte Carlo* (MCMC) methods [81, 82, 83, 84, 85], to sample the *transition path ensemble* (TPE) of Itô diffusion equations. The resulting algorithm, known as the *pCN* algorithm, is an MCMC procedure which can evaluate probabilistic path-integrals over the space of stochastic paths. Using Kosambi-Karhunen-Loève (KKL) theory [86, 87, 88], we develop a fast implementation of the pCN using Fast-Fourier Transforms. In Sec. 2.3.4 we investigate the band structure of stochastic paths in the KKL expansion. We provide both analytical and numerical evidence that general Itô diffusions with additive noise undergo a band separation in KKL mode-space, and that only the statistics of the lower band

is non-trivial, whilst the higher band is indistinguishable from that of free diffusion. In Sec. 2.3.5 we exploit the band structure of Itô diffusions to construct adapted step-sizes in the pCN method, using which we improved autocorrelation times. In Sec. 2.3.6 we provide a proof-of-concept of how to sample transition path ensembles of trajectories with generalised end-point conditions.

Further developing the algorithm, in Sec. 2.4 we present the *teleporter MCMC* (TMC), an extension of the pCN geared towards effectively sampling TPEs with multiple competing transition channels. We use semi-classical expansions [89, 90, 91, 92, 93, 94] of the path-probability measure to approximate the distribution over the transition path ensemble around local instantons, and combine these local Gaussian distributions with the pCN algorithm to allow for the Markov chain to ‘teleport’ between segregated regions of high probability in path-space. Similar schemes have been proposed previously in [95], where the chain can teleport between local uniform distributions around the the maxima of the target measure, and in [96] where a sequence Markov chains are run simultaneously. Our main innovation is that our algorithm is our use of the semi-classical Gaussian approximations, and that the algorithm is defined directly on the infinite-dimensional space of stochastic trajectories.

Other well-known techniques of sampling the TPE are the *transition path sampling* (TPS) [5, 97, 98, 99, 100] and *forward flux sampling* (FFS) [101, 102, 103]. One of the main differences in our proposed method is that it is defined for general Itô diffusions with additive noise, whilst established methods like the TPE and FFS tend to require that the system drift and noise satisfy the detailed balance condition, such that the dynamics is reversible. As our method is defined for arbitrary drift-fields, it can sample the TPEs of both equilibrium and non-equilibrium systems. Another point of difference is that we parametrise stochastic paths in a global basis, where the TPE and FFS use a uniform discretisation of stochastic paths. Specifically, we expand the stochastic paths of Itô diffusions in the Kosambi-Karhunen-Loève basis of Brownian bridge processes. As will be discussed in Sec. 2.3, we find that the statistics of the TPE separates into distinct low- and high-frequency bands in the KKL basis, where the higher band is completely decorrelated and behave is near that of a Brownian process. We find that the non-trivial statistics of the TPE is wholly contained in the lower band. Furthermore, our choice of basis also allows great freedom in the choice of initial- and final-conditions of the TPE. As for the TPS and FFS, we can choose paths to start and end in open regions in state-space. However, the pCN and TMC also admits fixed start- and end-point configurations. Perhaps one of the main differences. The TMC, which will be discussed in Sec. 2.4, also allows for the simultaneous sampling of multiple transition channels in the TPE simultaneously. Without recourse to techniques such as [104, 105], which requires a costly hierarchical sequence of MCMC algorithms running at different temperatures, established methods and traditional

MCMC methods will often fail to sample all transition channels. This is referred to as the *slow mixing* [106] of the modes of target probability distribution. Finally, another point of difference is that the pCN and TMC algorithms are mathematically formulated directly on the infinite-dimensional space of stochastic paths, whilst the TPE and FFS are defined on discretised path-spaces that approximate the true system under refinement of the discretisation-mesh.

In the next section, we will offer a brief review of the necessary mathematical theory for path-space MCMC methods. This will simultaneously also serve as an introduction to, and discussion of, the transition path ensemble and path-integral approaches to study this infinite-dimensional space of stochastic paths. In particular we aim to show that path-integrals, which are often described as the ‘formal limit’ of a finite-dimensional discretisation [29, 92, 107], can be understood without recourse to discretisation. Furthermore, in Sec. 2.3 we will see that this mathematical rigour is the necessary theoretical understanding that enables the construction of path-space MCMC methods. We will also discuss the distinctions between two common expressions for the path-integral approach: the Freidlin-Wentzell and the *Onsager-Machlup* path-probability densities, of which the former was discussed in Ch. 1. The main purpose of the treatment below is to explain the relevant mathematical concepts, as well as to recontextualise concepts familiar to the mathematical community to a physics audience. We will therefore eschew rigour whenever it facilitates more pedagogic explanations. The reader can see the references for more details.

2.2 Path-measures and densities for stochastic processes

Physically, the broad class of systems we consider are described by *overdamped Langevin equations*, which are often written as [19, 21]

$$\dot{\mathbf{X}} = \mu \mathbf{F}(\mathbf{X}) + \sqrt{2D} \boldsymbol{\xi} \quad (2.1)$$

where $\mathbf{X}(t)$ is a vector in \mathbb{R}^d representing the state of the system at time t , d is the dimension of the system, $\mathbf{F} : \mathbb{R}^d \rightarrow \mathbb{R}$ is a force-field on the system, $\mu \in \mathbb{R}^{d \times d}$ is the mobility, $\boldsymbol{\xi}(t) \in \mathbb{R}^d$ is a white-noise realisation with correlation function

$$\langle \xi_i(t) \xi_j(t') \rangle = 2D\delta_{ij}\delta(t - t') \quad (2.2)$$

and where D is the *diffusion matrix*. The diffusion constant is often written as $D = \mu\theta$, where θ is the temperature, in which case the prefactor of ξ in Eq. 2.1 is a matrix square-root. In general the diffusion matrix can be state-dependent $D(\mathbf{X})$, in which case we call

the noise *multiplicative*, and if it is not state-dependent we call the noise *additive*. In the following we will focus on systems with additive noise. We will also assume D is diagonal with constant entries, as this can always be achieved with coordinate transformations. This is equivalent to assuming that the mobility μ is a scalar constant.

The notation in Eq. 2.1 is highly formal, as the time-derivative of a stochastic process \mathbf{X} is not well-defined, and realisations of a white-noise process ξ are not functions, but distributions. However, the Langevin equation is equivalent to the Itô diffusion equation with additive noise

$$d\mathbf{X} = \mathbf{b}(\mathbf{X})dt + \sqrt{2D}d\mathbf{W} \quad (2.3)$$

which represents the stochastic displacement $d\mathbf{X}$ in a time interval dt , subject to a drift-field \mathbf{b} and Brownian displacements $\sigma d\mathbf{W}$, where \mathbf{W} is the Wiener process and $\sigma \in \mathbb{R}^{d \times d}$ is the *volatility matrix*. Equation 2.3 and Eq. 2.1 can be related by setting $\mathbf{b} = \mu\mathbf{F}$ and $\sigma = \sqrt{2D}$. Formally, we may relate \mathbf{W} and ξ as $\dot{\mathbf{W}} = \xi$. Equation 2.3 can be understood as a short-hand for an alternative integral representation of the Itô diffusion equation

$$\mathbf{X}(t) = \int_0^t \mathbf{b}(\mathbf{X}(t'))dt' + \sqrt{2D} \int_0^t d\mathbf{W}(t') \quad (2.4)$$

where the second term is an *Itô stochastic integral* [55, 56] with respect to the stochastic increment $d\mathbf{W}$.

Let $C^0([0, T])$ be the set of continuous paths $\mathbf{x} : [0, T] \rightarrow \mathbb{R}^d$, and furthermore let $C_{\mathbf{x}_0}^0([0, T]) \subset C^0([0, T])$ be the subset of paths that satisfy $\mathbf{x}(0) = \mathbf{x}_0$, where $\mathbf{x}_0 \in \mathbb{R}^d$ is the star-point of the path. The *transition path ensemble* $E_{\mathbf{x}_0}^{\mathbf{x}_T}([0, T]) \subset C_{\mathbf{x}_0}^0([0, T])$ of Eq. 2.3 is the subset of paths that satisfy $\mathbf{x}(0) = \mathbf{x}_0$ and $\mathbf{x}(T) = \mathbf{x}_T$, where $\mathbf{x}_T \in \mathbb{R}^d$ is the end-point of the path. Realisations of Eq. 2.3 are elements $\mathbf{X} \in C_{\mathbf{x}_0}^0([0, T])$, and we will often refer to these as *stochastic paths*. Given an end-point \mathbf{x}_T , a *transition path* is a stochastic path that is an element $\mathbf{X} \in E_{\mathbf{x}_0}^{\mathbf{x}_T}([0, T])$. Transition paths can thus be considered realisations of a *pinned* Itô diffusion, where we condition on the end-point $\mathbf{X}(T) = \mathbf{x}_T$.

Consider some measurable subset of stochastic paths $U \subset C_{\mathbf{x}_0}^0([0, T])$. The *law* $\mathbb{P}_{\mathbf{X}}$ of a stochastic differential equation like Eq. 2.3 is a *probability measure* over $C_{\mathbf{x}_0}^0([0, T])$. The probability of observing a stochastic trajectory in a measurable subset $U \subset C_{\mathbf{x}_0}^0([0, T])$ is then $\mathbb{P}_{\mathbf{X}}(U)$. We will refer to $\mathbb{P}_{\mathbf{X}}$ as both a *law* and a *path measure* interchangeably. In the physics literature $\mathbb{P}_{\mathbf{X}}(U)$ is often written in a path-integral formulation as [89, 108, 109]

$$\mathbb{P}_{\mathbf{X}}(U) = \int_U P_{\mathbf{X}}[\mathbf{x}] \mathcal{D}\mathbf{x} \quad (2.5)$$

where

$$P_{\mathbf{X}}[\mathbf{x}] \propto \exp(-S[\mathbf{x}]) \quad (2.6)$$

represents a *path-space probability density* over $C_{\mathbf{x}_0}^0([0, T])$, and the exponential scaling-factor of the density $S[\mathbf{x}]$ is called the *stochastic action*. In Eq. 2.5 we expressed the law $\mathbb{P}_{\mathbf{x}}$ in terms of the *path-probability density* $P_{\mathbf{x}}$. $\mathcal{D}\mathbf{x}$ represents what would be a volume measure over the infinite-dimensional space of stochastic paths. Mathematically, one would call $\mathcal{D}\mathbf{x}$ an infinite-dimensional analogue of a *Lebesgue measure* (although as we will discuss below, such an analogue does not exist). Equation 2.5 applies equivalently to subsets $A \subset E_{\mathbf{x}_0}^{\mathbf{x}_T}([0, T])$, by simply conditioning Eq. 2.6 on the end-point. Accordingly, we can compute the expectations of observables as

$$\langle \mathcal{O}[\mathbf{x}] \rangle = \int_{C_{\mathbf{x}_0}^0} \mathcal{O}[\mathbf{x}] P_{\mathbf{x}}[\mathbf{x}] \mathcal{D}\mathbf{x} \quad (2.7)$$

where $\mathcal{O}[\mathbf{x}]$ is a functional of stochastic paths $\mathbf{x} \in C_{\mathbf{x}_0}^0$ representing an observable. Equivalently, we can compute the expectations of observables in TPE by computing the integral over $E_{\mathbf{x}_0}^{\mathbf{x}_T}([0, T])$.

There are two well-known expressions for the action. The *Freidlin-Wentzell* (FW) action functional is [2, 58, 109, 110]

$$S_{\text{FW}}[\mathbf{x}] = \int_0^T \frac{1}{2D} |\dot{\mathbf{x}} - \mathbf{b}(\mathbf{x})|^2 dt \quad (2.8)$$

and the *Onsager-Machlup* (OM) action functional is [25, 26, 109, 111, 112, 113, 114]

$$S_{\text{OM}}[\mathbf{x}] = \int_0^T \left\{ \frac{1}{2D} |\dot{\mathbf{x}} - \mathbf{b}(\mathbf{x})|^2 + \frac{1}{2} \nabla \cdot \mathbf{b}(\mathbf{x}) \right\} dt. \quad (2.9)$$

Note that Eq. 2.8 is not interpreted here as a large deviation principle in a zero-temperature limit, but rather defines a path-probability density via Eq. 2.6 for arbitrary temperatures. This seeming contradiction, along with the fact that there are two stochastic actions, yielding ostensibly two different path-probability densities, has been a point of confusion [32, 109]. In [109] it is shown that the actions are equivalent for the purposes of sampling the path-probability distribution. In other words, formally Eq. 2.5 yields the same value for the probability $\mathbb{P}_{\mathbf{x}}(A)$ regardless of what action is used. However, in [32] it is shown that if one considers the subset $C_{\mathbf{x}_0}^1([0, T]) \subset C_{\mathbf{x}_0}^0([0, T])$ of paths with at-least one derivative, then the appropriate action functional is the Onsager-Machlup action. Therefore if one wants to consider *most-probable paths*, which are always piece-wise smooth, the Onsager-Machlup action is to be preferred over the Freidlin-Wentzell action. Note that in the limit of vanishing noise, which is the regime in which we worked in Ch. 1, the Onsager-Machlup action converges to the Freidlin-Wentzell action.

In terms of path-probability densities, there is in fact an infinite family of equally valid stochastic actions [32]. This is due to the right-hand side of Eq. 2.5 being ill-defined unless

an explicit discretisation scheme is prescribed

$$\mathbf{x} \longrightarrow \mathbf{x}_i = \mathbf{x}(i\Delta t), \quad (2.10)$$

such that

$$\dot{\mathbf{x}} \longrightarrow \frac{\mathbf{x}_{i+1} - \mathbf{x}_i}{\Delta t}, \quad (2.11)$$

where $i = 0, 1, \dots, N_T$ and δt is the time-step of the discretisation. Under such a discretisation the path-space volume density is replaced with

$$\mathcal{D}\mathbf{x} \longrightarrow \prod_{i=0}^{N_T} d\mathbf{x}_i. \quad (2.12)$$

We thus replace the path-space probability density $P_{\mathbf{X}}[\mathbf{X}]$ with a finite-dimensional probability density $p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_T})$. The discretised probability distribution approximates the statistics of the true process \mathbf{X} , and this approximation is improved with increasing N_T . However, we should note that the limit $N_T \rightarrow \infty$ does not exist. That Eq. 2.5 is not well-defined on the space continuous paths can be seen by evaluating any of the actions on an un-differentiable stochastic path, which then diverge due to the $\frac{1}{2D}|\dot{\mathbf{x}}|^2$ term. Eq. 2.5 should therefore only be interpreted as a formal expression for the path-probability.

Mathematically, the root of the problem stems from the non-existence of the ‘path-space volume measure’ $\mathcal{D}\mathbf{x}$. Consider some probability measure μ with density p over a finite-dimensional space \mathbb{R}^n . Probabilities can then be computed using integrals $\mu(U) = \int_U p(\mathbf{z})d\mathbf{z}$ for a measurable subset $U \subset \mathbb{R}^n$. Here, $d\mathbf{z}$ is a volume measure on the space \mathbb{R}^d , and is known in the mathematical literature as a *Lebesgue measure*. In particular, Lebesgue measures have the desirable property of translation-invariance.¹ Intuitively we can see that $\mathcal{D}\mathbf{x}$ formally plays the role of an infinite-dimensional Lebesgue measure from Eq. 2.12. However, it is a theorem that there exists no infinite-dimensional analogue of a Lebesgue measure [111, 115].

Although not a Lebesgue measure, there is a measure that is well-defined on general infinite dimensional spaces, known as the *abstract Wiener space construction* [115, 116]. If the sample space is $C_{\mathbf{x}_0}^0([0, T])$, then the corresponding probability measure $\mathbb{P}_{\mathbf{W}}$ is called the *classical Wiener measure* [117, 118]. As its name suggests, $\mathbb{P}_{\mathbf{W}}$ is the law of the Wiener process. If we consider the Wiener process as an Itô diffusion with the drift-field set to zero $\mathbf{b} = 0$, we can write the Wiener measure using Eqs. 2.5, 2.6 and 2.8 (or Eq. 2.9) as

$$\mathbb{P}_{\mathbf{W}}(U) = \int_U P_{\mathbf{W}}[\mathbf{x}] \mathcal{D}\mathbf{x} \quad (2.13)$$

¹For example, the measure $d\mathbf{z}$ is translation-invariant under uniform coordinate transformations $\mathbf{z} \rightarrow \mathbf{z} + \mathbf{v}$, where $\mathbf{v} \in \mathbb{R}^d$ is some constant vector.

where $A \subset C_{\mathbf{x}_0}^0([0, T])$ and

$$P_{\mathbf{W}}[\mathbf{x}] \propto \exp(-S_0[\mathbf{x}]) \quad (2.14a)$$

$$S_0[\mathbf{x}] = \int_0^T \frac{1}{2D} |\dot{\mathbf{x}}|^2 dt \quad (2.14b)$$

with the understanding that Eq. 2.13 is to be interpreted formally in the sense described above. Under any finite-dimensional discretisation, the resulting probability density because a multi-variate Gaussian. We thus call $\mathbb{P}_{\mathbf{W}}$ a *Gaussian measure*, and indeed \mathbf{W} is an example of a Gaussian process.

Formally, we can now decompose the path-space probability density Eq. 2.6 as

$$\begin{aligned} P_{\mathbf{X}}[\mathbf{x}] &= (P_{\mathbf{X}}/P_{\mathbf{W}})P_{\mathbf{W}} \\ &= \frac{d\mathbb{P}_{\mathbf{X}}}{d\mathbb{P}_{\mathbf{W}}}[\mathbf{x}]P_{\mathbf{W}}[\mathbf{x}] \end{aligned} \quad (2.15)$$

where we have identified $P_{\mathbf{X}}/P_{\mathbf{W}}$ as the *Radon-Nikodym derivative* of $\mathbb{P}_{\mathbf{X}}$ with respect to $\mathbb{P}_{\mathbf{W}}$. We find

$$\frac{d\mathbb{P}_{\mathbf{X}}}{d\mathbb{P}_{\mathbf{W}}}[\mathbf{x}] \propto \exp(-\Phi[\mathbf{x}]) \quad (2.16)$$

and $\Phi[\mathbf{x}] = S_{\mathbf{FW}}[\mathbf{x}] - S_0[\mathbf{x}]$ is the *relative action*, which is equivalently $\Phi[\mathbf{x}] = S_{\mathbf{OM}}[\mathbf{x}] - S_0[\mathbf{x}]$. Intuitively, we can understand $\frac{d\mathbb{P}_{\mathbf{X}}}{d\mathbb{P}_{\mathbf{W}}}$ as the probability density of $\mathbb{P}_{\mathbf{X}}$ with respect to the Wiener measure $\mathbb{P}_{\mathbf{W}}$. Compare this with $P_{\mathbf{X}}$, which we formally understand to be a probability density with respect to the fictitious path-space volume measure $\mathcal{D}\mathbf{x}$. In other words, $\frac{d\mathbb{P}_{\mathbf{X}}}{d\mathbb{P}_{\mathbf{W}}}$ can be seen as a re-weighting of the probability distribution over stochastic Wiener realisations \mathbf{W} to the realisations \mathbf{X} of the Itô equation, whilst $P_{\mathbf{X}}$ re-weights the ‘uniform distribution’ over the space of continuous paths.

We now introduce some notation which will be of use in the subsequent section. For any two probability measures \mathbb{P} and \mathbb{P}_0 defined on a sample space C , if the Radon-Nikodym derivative $\frac{d\mathbb{P}}{d\mathbb{P}_0}$ exists, we say that \mathbb{P} has a *density with respect to* \mathbb{P}_0 . Furthermore, the existence of $\frac{d\mathbb{P}}{d\mathbb{P}_0}$ implies that \mathbb{P} is *absolutely continuous* with respect to \mathbb{P}_0 , which means that

$$\mathbb{P}_0(A) = 0 \Rightarrow \mathbb{P}(A) = 0 \quad (2.17)$$

for all measurable subsets $A \subset C$, where \Rightarrow signifies implication. We write this as $\mathbb{P} \ll \mathbb{P}_0$.

As opposed to $P_{\mathbf{X}}$, the density $\frac{d\mathbb{P}_{\mathbf{X}}}{d\mathbb{P}_{\mathbf{W}}}$ can be evaluated on un-differentiable stochastic

paths. To see this, consider

$$\begin{aligned}\Phi[\mathbf{x}] &= S_{\text{FW}}[\mathbf{x}] - S_0[\mathbf{x}] \\ &= \int_0^T \left\{ \frac{1}{2D} |\mathbf{b}(\mathbf{x})|^2 - \frac{1}{D} \mathbf{b}(\mathbf{x}) \cdot \dot{\mathbf{x}} + \right\} dt \\ &= \int_0^T \frac{1}{2D} |\mathbf{b}(\mathbf{x})|^2 dt - \int_{\mathbf{x}_0}^{\mathbf{x}_T} \frac{1}{D} \mathbf{b}(\mathbf{x}) \cdot d\mathbf{x}.\end{aligned}\tag{2.18}$$

The integrals in the third line are well-defined for un-differentiable paths. Therefore, the second term of the relative action becomes a stochastic integral if we evaluate it on a realisation of the Itô process

$$\Phi[\mathbf{X}] = \int_0^T \frac{1}{2D} |\mathbf{b}(\mathbf{X})|^2 dt - \int_{\mathbf{x}_0}^{\mathbf{x}_T} \frac{1}{D} \mathbf{b}(\mathbf{X}) \cdot d\mathbf{X}.\tag{2.19}$$

Like any integral Eq. 2.19 has to be discretised in order to be evaluated numerically. However, in contrast to the finite-dimensional approximation $p(\mathbf{x}_1, \dots, \mathbf{x}_{N_T})$ of the path-probability density $P[\mathbf{x}]$, the discretisation of Eq. 2.19 has a well-defined limit at $N_T \rightarrow \infty$. As before, both the Freidlin-Wentzell and Onsager-Machlup forms of Φ are statistically equivalent. Furthermore, as showed in [32], an infinite family of expressions of equivalent Φ , which include the OM and FW forms, can be derived from Eq. 2.19 depending on the discretisation scheme.

We can now rewrite Eq. 2.5 in terms of well-defined quantities. Since

$$\mathbb{P}_{\mathbf{X}}(U) = \int_A P_{\mathbf{X}}[\mathbf{x}] \mathcal{D}\mathbf{x} = \int_U \frac{d\mathbb{P}_{\mathbf{X}}}{d\mathbb{P}_{\mathbf{W}}}[\mathbf{x}] P_{\mathbf{W}}[\mathbf{x}] \mathcal{D}\mathbf{x}\tag{2.20}$$

we get

$$\mathbb{P}_{\mathbf{X}}(U) = \int_U \frac{d\mathbb{P}_{\mathbf{X}}}{d\mathbb{P}_{\mathbf{W}}}[\mathbf{x}] d\mathbb{P}_{\mathbf{W}}[\mathbf{x}]\tag{2.21}$$

where we have identified $P_{\mathbf{W}}[\mathbf{x}] \mathcal{D}\mathbf{x} = d\mathbb{P}_{\mathbf{W}}[\mathbf{x}]$. Although Eq. 2.21 was derived via our ill-defined formal path-integral expressions, the resulting expression we have arrived at is nevertheless correct [82, 83, 84].

Just as $\mathbb{P}_{\mathbf{W}}$ serves as a measure over the space $C_{\mathbf{x}_0}([0, T])$ of stochastic paths, we can similarly find a well-defined measure over the transition path ensemble $E_{\mathbf{x}_0}^{\mathbf{x}_T}([0, T])$. Let \mathbf{B} be the *Brownian bridge process*, which can be defined as a Wiener process conditioned as $\mathbf{B}(0) = \mathbf{x}_0$ and $\mathbf{B}(T) = \mathbf{x}_T$. In other words, we can write it as

$$\mathbf{B} = \left(1 - \frac{t}{T}\right) \mathbf{W} + \frac{t}{T} (\mathbf{x}_T - \mathbf{x}_0)\tag{2.22}$$

if we let $\mathbf{W}(0) = \mathbf{x}_0$. Then the law $\mathbb{P}_{\mathbf{B}}$ is defined over $E_{\mathbf{x}_0}^{\mathbf{x}_T}([0, T])$, and we have that [82]

$$\mathbb{P}_{\mathbf{X}}(U) = \int_U \frac{d\mathbb{P}_{\mathbf{X}}}{d\mathbb{P}_{\mathbf{B}}}[\mathbf{x}] d\mathbb{P}_{\mathbf{B}}[\mathbf{x}] \quad (2.23)$$

where $\mathbb{P}_{\mathbf{X}}$ is here to be understood as the law of the pinned Itô diffusion, where realisations \mathbf{X} are conditioned as $\mathbf{X}(0) = \mathbf{x}_0$ and $\mathbf{X}(T) = \mathbf{x}_T$. Note that we have used the same symbol $\mathbb{P}_{\mathbf{X}}$ for the law of both the unpinned and pinned processes. Henceforth, $\mathbb{P}_{\mathbf{X}}$ will denote to law of latter, unless explicitly stated otherwise.

Despite the fact that densities over path-spaces are ill-defined, we will continue making use of the path-integrals with respect to the fictitious infinite-dimensional Lebesgue measure. This is primarily due to the ubiquity of the path-integral in the physics, and because it is often intuitively easier to understand densities with respect to a uniform volume density. Mathematically, this will not cause any issues as long as we understand that ill-defined path-integrals correspond to well-defined integrals with respect to the Wiener measure, as in Eq. 2.21 and Eq. 2.23.

This concludes our mathematical review of path measures and densities. The above discussion can be seen as an interpolation between the physicist's and the mathematician's understanding of probability. The former is familiar with the notion of a probability *density* $p(\mathbf{x})$, defined on a sample space C , where $\mathbf{x} \in C$ represents a possible outcome. On the other hand, in the mathematics literature probabilities are as frequently described in terms of probability *measure*; functions between measurable subsets of C into $[0, 1]$. It is a theorem that in the infinite-dimensional setting, there is no well-defined notion of probability density analogous to that of the finite-dimensional setting. Despite this fact, path-space densities are ubiquitous in physics. We showed how the notion of a path-space density can be understood, and how they can be related to well-defined infinite-dimensional measures. The above discussion thus has some utility beyond outlining the necessary mathematical exposition for the results of the subsequent sections. Namely, our aim was to show that formal path-integral expressions can be used with impunity, as long as they can be rooted in terms of well-defined path measures.

2.3 Path-space MCMC methods

In this section we will discuss and apply the *preconditioned Nicolson-Crank* (pCN) algorithm to sample the transition path ensemble of Itô diffusions with additive noise. In Sec. 2.3.1 we begin with an overview of Markov chain Monte Carlo methods, and proceed to reproduce the derivation of the general pCN algorithm in [81, 82, 83, 84, 85]. We will frequently express probabilities in terms of fictitious path-space probability densities with respect to Lebesgue measures, as this is the predominant mathematical language used in the

physics literature. The reader may refer to the references for a more rigorous approach. In Sec. 2.3.2 we expand stochastic paths in the Kosambi-Karhunen-Lo  ve (KKL) basis of the Brownian bridge process, and develop a sampling algorithm using Fast-Fourier Transforms. In Sec. 2.3.4 we analyse the band structure of It   diffusions through the lens of the KKL basis, and in Sec. 2.3.5 we exploit the band-structure to improve the autocorrelation times of the MCMC algorithm. The results of this section will lead up to and be further developed in Sec. 2.4, where we extend the pCN algorithm to simultaneously sample multiple transition channels.

2.3.1 The preconditioned Crank-Nicolson algorithm

Markov chain Monte Carlo (MCMC) methods refers to a general class of methods of sampling a target probability measure \mathbb{P} , wherein one constructs a Markov chain that has \mathbb{P} as its steady-state measure. \mathbb{P} is equivalently called the *target measure* of the MCMC protocol, and the *invariant measure* of the Markov chain. The key idea is that as the Markov chain is simulated, each state in the chain will asymptotically sample \mathbb{P} as we simulate for sufficiently many steps.

Here will consider MCMC methods to sample a general target measure \mathbb{P} with sample space C . We will begin by first discussing general MCMC procedures, and then proceed to define the precondition Crank-Nicolson algorithm. Let P be the probability density of \mathbb{P} with respect to the Lebesgue measure, and we write this relation as $P \sim \mathbb{P}$. We will in general consider C to be an infinite-dimensional functional space, in which case P is a fictitious density in the sense discussed in the introduction of this chapter. We also assume that \mathbb{P} is absolutely continuous with respect to a measure \mathbb{P}_0 , such that we can write

$$\mathbb{P}(U) = \int_U \frac{d\mathbb{P}}{d\mathbb{P}_0}[\mathbf{x}] d\mathbb{P}_0[\mathbf{x}] \quad (2.24)$$

for any measurable subset $A \subset C$. We furthermore assume that \mathbb{P}_0 is a Gaussian measure, and we refer to it as the Gaussian *reference measure* of \mathbb{P} . We express the densities $P_0 \sim \mathbb{P}_0$ and $P \sim \mathbb{P}$ as

$$P_0[\mathbf{x}] \propto \exp(-S_0[\mathbf{x}]) \quad (2.25a)$$

$$P[\mathbf{x}] \propto \exp(-S_0[\mathbf{x}] - \Phi[\mathbf{x}]) = P_0[\mathbf{x}] \exp(-\Phi[\mathbf{x}]) \quad (2.25b)$$

where $S_0[\mathbf{x}]$ is the Gaussian action of $P_0[\mathbf{x}]$ and $\Phi[\mathbf{x}]$ is the relative action, which satisfies $\frac{d\mathbb{P}}{d\mathbb{P}_0} \propto \exp(-\Phi[\mathbf{x}])$. In subsequent subsections we will apply the MCMC method to sample the probability measure $\mathbb{P}_{\mathbf{x}}$ over the sample space $E_{\mathbf{x}_0}^{\mathbf{x}_T}([0, T])$, which has a density with respect to the Gaussian measure $\mathbb{P}_{\mathbf{B}}$.

The aim of the MCMC is to generate samples from the target measure \mathbb{P} . In applications,

this is often for the purpose of estimating expectations like Eq. 2.7. Let $\mathbf{x}^{(n)}$, $n = 1, \dots, M$ be M samples from the target measure \mathbb{P} . Then expectations can be approximated as

$$\langle \mathcal{O}[\mathbf{x}] \rangle = \int_C \mathcal{O}[\mathbf{x}] \frac{d\mathbb{P}}{d\mathbb{P}_0}[\mathbf{x}] d\mathbb{P}_0[\mathbf{x}^{(n)}] \approx \frac{1}{M} \sum_{n=1}^M \mathcal{O}[\mathbf{x}]. \quad (2.26)$$

for any observable $\mathcal{O}[\mathbf{x}]$. In the limit $M \rightarrow \infty$, the right-hand side will converge to the true expectation.

Consider a Markov chain with states $\mathbf{x}^{(n)} \in C$, where C is the sample space of a target measure \mathbb{P} . If $\mathbf{x}^{(n)} \in E$ is the current state of the Markov chain, its next step $\mathbf{x}^{(n+1)}$ is drawn from the *transition kernel* $T[\mathbf{x}^{(n)} \rightarrow \mathbf{x}_{i+1}]$, which is a probability distribution in $\mathbf{x}^{(n+1)} \in C$. If $\mathbb{P}_{\mathbf{X}}$ is the invariant measure of the Markov chain then the transition kernel satisfies

$$\int_C P[\mathbf{x}] T[\mathbf{x} \rightarrow \mathbf{y}] D\mathbf{x} = P[\mathbf{y}]. \quad (2.27)$$

For any given target measure \mathbb{P} there is an infinite family of possible Markov chains, defined by a transition kernel $T[\mathbf{x} \rightarrow \mathbf{y}]$ that has \mathbb{P} as its invariant measure. The particular prescription by which $T[\mathbf{x} \rightarrow \mathbf{y}]$ is constructed given \mathbb{P} is what defines the MCMC.

A popular MCMC variant is the *Metropolis-Hastings scheme*[119, 120], which defines the Markov chain in terms of a *proposal kernel* $Q[\mathbf{x} \rightarrow \mathbf{y}]$ and an *acceptance kernel* $A[\mathbf{x}, \mathbf{y}]$. These are constructed in such a way that the transition kernel satisfies

$$P[\mathbf{x}] T[\mathbf{x} \rightarrow \mathbf{y}] = P[\mathbf{y}] T[\mathbf{y} \rightarrow \mathbf{x}]. \quad (2.28)$$

which in physics is known as a *detailed balance* condition on the Markov chain with respect to \mathbb{P} . The transition kernel is written in terms of the proposal and acceptance kernels as

$$T[\mathbf{x} \rightarrow \mathbf{y}] = Q[\mathbf{x} \rightarrow \mathbf{y}] A[\mathbf{x}, \mathbf{y}] + \delta[\mathbf{x} - \mathbf{y}] \int_C (1 - A[\mathbf{x}, \mathbf{z}]) Q[\mathbf{x} \rightarrow \mathbf{z}] D\mathbf{z}. \quad (2.29)$$

where $\delta[\mathbf{x}]$ represents a Dirac-delta distribution on functional space. Formally, we may write

$$\delta[\mathbf{x}] = \begin{cases} \infty, & \text{if } \mathbf{x} = 0 \\ 0, & \text{otherwise.} \end{cases} \quad (2.30)$$

By substituting Eq. 2.29 into Eq. 2.28 we find the following expression for the acceptance probability

$$A[\mathbf{x}, \mathbf{y}] = \min \left\{ 1, \frac{V^T[\mathbf{x}, \mathbf{y}]}{V[\mathbf{x}, \mathbf{y}]} \right\} \quad (2.31)$$

where

$$V[\mathbf{x}, \mathbf{y}] = P[\mathbf{x}]Q[\mathbf{x} \rightarrow \mathbf{y}]. \quad (2.32a)$$

$$V^T[\mathbf{x}, \mathbf{y}] = P[\mathbf{y}]Q[\mathbf{y} \rightarrow \mathbf{x}] \quad (2.32b)$$

are probability densities over $(\mathbf{x}, \mathbf{y}) \in C \times C$ with respect to the measures $\mathbb{V} \sim V$ and $\mathbb{V}^T \sim V^T$, and we will refer to them as the forward and backwards kernels respectively. In general Eq. 2.31 is only well-defined if V^T vanishes whenever V vanishes. In other words, we must have that $\mathbb{V}^T \ll \mathbb{V}$, in which case we can rewrite Eq. 2.31 as

$$A[\mathbf{x}, \mathbf{y}] = \min \left\{ 1, \frac{d\mathbb{V}^T}{d\mathbb{V}}[\mathbf{x}, \mathbf{y}] \right\}. \quad (2.33)$$

We see that the MCMC protocol is only well-defined if the proposal kernel $Q[\mathbf{x} \rightarrow \mathbf{y}]$ is such that the forward and backward kernels satisfy the absolute continuity condition $\mathbb{V}^T \ll \mathbb{V}$.

Rather than by trying to parse Eq. 2.29, the Metropolis-Hastings algorithm can be more easily understood by summarising it algorithmically as follows:

1. Choose an initial state $\mathbf{x}^{(0)} \in E$.
2. Draw a proposal state \mathbf{y} from the distribution $Q[\mathbf{x}^{(n)} \rightarrow \mathbf{y}]$.
3. Compute the *acceptance probability* $A[\mathbf{x}^{(n)}, \mathbf{y}]$.
4. Draw a random number V from $\text{Unif}([0, 1])$.
 - If $V < A[\mathbf{x}^{(n)}, \mathbf{y}]$ then set $\mathbf{x}^{(n+1)} = \mathbf{y}$.
 - Otherwise set $\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)}$.
5. Repeat step 2.

where $\text{Unif}([0, 1])$ is the uniform distribution over the unit interval. After M steps, we have a chain of states $L_M = \{\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}\}$. Asymptotically, as $M \rightarrow \infty$, the set L_M will converge towards a sample of the target distribution \mathbb{P} . The rate of convergence will in general depend on the implementation of the Metropolis-Hastings rule.

The remaining degree of freedom in the Metropolis-Hastings algorithm, which is to choose the proposal kernel $Q[\mathbf{x} \rightarrow \mathbf{y}]$. A common choice is the update rule

$$\mathbf{y} = \mathbf{x} + \kappa \mathbf{w} \quad (2.34)$$

where κ is a step-size parameter, \mathbf{x} is the current state of the Markov chain, \mathbf{y} is the proposal state, and \mathbf{w} is a sample drawn from the Gaussian distribution \mathbb{P}_0 . The corresponding

kernel density can be written in terms of the density $P_0 \sim \mathbb{P}_0$ of the reference measure as

$$Q_{\text{RMW}}[\mathbf{x} \rightarrow \mathbf{y}] \propto \exp(-S_0[\mathbf{x} - \mathbf{y}]/\kappa^2) \quad (2.35)$$

Eq. 2.34 defines the *Gaussian random-walk Metropolis-Hastings scheme* (RMW) [119, 120].

In practice to use the RMW in numerics, or any MCMC, we must prescribe some discretisation procedure on the functional space C . In general this will mean parameterising sample space elements as $\mathbf{x} = \mathbf{x}(\mathbf{a}) \in C$, where $\mathbf{a} \in \mathbb{R}^N$ and N is the discretisation order. Effectively, this induces a target measure \mathbb{P}^N on the sample space

$$C^N = \{\mathbf{x}(\mathbf{a}) : \mathbf{a} \in \mathbb{R}^N\} \subset C \quad (2.36)$$

which is a finite-dimensional subset. Accordingly, the proposal and acceptance kernels now become densities over finite-dimensional distributions $Q(\mathbf{a} \rightarrow \mathbf{b})$ and $A(\mathbf{a}, \mathbf{b})$ respectively. For any finite N , the resulting MCMC algorithm is well-defined and will asymptotically sample \mathbb{P}^N when running the Markov chain for sufficiently many steps. Furthermore, the larger the N , the better will the discretised target measure \mathbb{P}^N approximate the true target measure \mathbb{P} . However, it is well-documented that standard MCMC schemes like the RMW suffer the curse of dimensionality [121] for infinite-dimensional systems [81, 85]. That is, as N is increased the algorithm suffers a curse of dimensionality, so that the number of steps required for the Markov chain to converge onto the target \mathbb{P}^N diverges with N .

The root cause of the divergence of the convergence-times of the RMW when increasing N is that the protocol is not well-defined in the $N \rightarrow \infty$ limit. In other words, the RMW is not a well-defined Markov chain on infinite-dimensional spaces. This can be seen by noting that the backwards kernel measure \mathbb{V}^T is not absolutely continuous with respect to the forward kernel measure \mathbb{V} . If $\mathbb{V}^T \ll \mathbb{V}$, then from Eq. 2.17 we would have that

$$\int_{U \times U'} V[\mathbf{x}, \mathbf{y}] \mathcal{D}\mathbf{x} \mathcal{D}\mathbf{y} = 0 \Rightarrow \int_{U \times U'} V^T[\mathbf{x}, \mathbf{y}] \mathcal{D}\mathbf{x} \mathcal{D}\mathbf{y} = 0 \quad (2.37)$$

where $U \times U' \subset C \times C$. However, if we substitute Eq. 2.35 into Eq. 2.37, using Eq. 2.32, we find that absolute continuity fails. Due to the symmetry of the RMW proposal kernel it is clear that Eq. 2.37 does not hold for arbitrary $U \times U'$.

Recent work by [81, 82, 83, 84] have developed and formalised the theory of MCMC methods so that protocols can remain well-defined on infinite-dimensional functional sample spaces. One of the main results of this work was the *preconditioned Crank-Nicolson* (pCN) algorithm. In contrast to the RMW, and most other common MCMC protocols, the pCN is defined directly on the space of functions. This guarantees that, upon discretisation, the algorithm remains well-defined under mesh refinement $N \rightarrow \infty$.

The pCN algorithm overcomes the degeneracy suffered by the RMW by imposing that

its proposal kernel $Q_{\text{pCN}}[\mathbf{x} \rightarrow \mathbf{y}]$ is in detailed balance with a Gaussian reference measure that the target measure is absolutely continuous with respect to. For example, as $\mathbb{P} \ll \mathbb{P}_0$ we may impose

$$P_0[\mathbf{x}]Q_{\text{pCN}}[\mathbf{x} \rightarrow \mathbf{y}] = P_0[\mathbf{y}]Q_{\text{pCN}}[\mathbf{y} \rightarrow \mathbf{x}] \quad (2.38)$$

which ensures that $\mathbb{V}^T \ll \mathbb{V}$. To see this, define the measures $\mathbb{V}_0 \sim P_0[\mathbf{x}]Q_{\text{pCN}}[\mathbf{x} \rightarrow \mathbf{y}]$ and $\mathbb{V}_0^T \sim P_0[\mathbf{y}]Q_{\text{pCN}}[\mathbf{y} \rightarrow \mathbf{x}]$, which from Eq. 2.38 clearly satisfy mutual absolute continuity. $\mathbb{V}^T \ll \mathbb{V}$ then follows from $\mathbb{P} \ll \mathbb{P}_0$, and using the transitivity of absolute continuity. The pCN proposal kernel is defined by

$$\mathbf{y} = \sqrt{1 - \kappa^2}\mathbf{x} + \kappa\mathbf{w} \quad (2.39)$$

where \mathbf{w} is drawn from \mathbb{P}_0 as before. In terms a density, the proposal kernel is

$$Q_{\text{pCN}}[\mathbf{x} \rightarrow \mathbf{y}] \propto \exp(-S_0[\sqrt{1 - \kappa^2}\mathbf{x} - \mathbf{y}]/\kappa^2) \quad (2.40)$$

which satisfies Eq. 2.38. From Eq. 2.31 and Eq. 2.32 we find that the resulting acceptance probability has the elegant expression

$$A_{\text{pCN}}[\mathbf{x}, \mathbf{y}] = \min \{1, \exp(\Phi[\mathbf{x}] - \Phi[\mathbf{y}])\}. \quad (2.41)$$

Now as opposed to $P[\mathbf{x}]$ and $P_0[\mathbf{x}]$, since $\Phi[\mathbf{x}]$ is the exponential factor of the well-defined density $\frac{d\mathbb{P}}{d\mathbb{P}_0}$ it must in-turn also be well-defined for all $\mathbf{x} \in C$. Therefore Eq. 2.41 is also well-defined.

We can now state the pCN algorithm in its entirety:

1. Choose an initial state $\mathbf{x}^{(0)}$.
2. Generate a proposal state

$$\mathbf{y} = \sqrt{1 - \kappa^2}\mathbf{x}^{(n)} + \kappa\mathbf{w} \quad (2.42)$$

where \mathbf{w} is drawn from the reference measure \mathbb{P}_0 .

3. Compute the acceptance probability $A_{\text{pCN}}[\mathbf{x}^{(n)}, \mathbf{y}]$.
4. Draw a random number V from $\text{Unif}([0, 1])$.
 - If $V < A_{\text{pCN}}[\mathbf{x}^{(n)}, \mathbf{y}]$ then set $\mathbf{x}^{(n+1)} = \mathbf{y}$.
 - Otherwise set $\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)}$.
5. Repeat step 2.

2.3.2 Kosambi-Karhunen-Loève expansions of stochastic trajectories

In the previous subsection we outlined the pCN algorithm, which was defined directly on general infinite-dimensional functional sample spaces. Here we specialise to transition paths of pinned Itô diffusions. The sample space is now $E_{\mathbf{x}_0}^{\mathbf{x}_T}([0, T])$, the transition path ensemble, and the target measure is the law $\mathbb{P}_{\mathbf{X}}$. Henceforth we will abbreviate the transition path ensemble as $E \equiv E_{\mathbf{x}_0}^{\mathbf{x}_T}([0, T])$. As shown in Eq. 2.23, $\mathbb{P}_{\mathbf{X}}$ has a density with respect to the Brownian bridge law $\mathbb{P}_{\mathbf{B}}$, and we will consider the latter the reference Gaussian measure.

In the second step of the pCN algorithm, we must sample the reference Gaussian measure to generate the proposal state. One method by which to sample $\mathbb{P}_{\mathbf{B}}$ is using the *Kosambi-Karhunen-Loève* (KKL) theorem [86, 87, 88], which is a representation of a stochastic process as an infinite linear combination of orthogonal basis functions $\phi_k(t)$. The KKL representation is found by solving a homogeneous Fredholm integral equation of the second kind

$$\int K_Y(s, t)\phi_i(s)ds = \lambda_i\phi_i(t) \quad (2.43)$$

where $K_Y(s, t)$ is the covariance function of a given, in this case one-dimensional, stochastic process Y . If Y is *centered* (in other words, if $\langle Y(t) \rangle = 0$ for all $t \in [0, T]$), then the KKL expansion of Y is

$$Y(t) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} Z_k \phi_k(t) \quad (2.44)$$

where $\phi_i(t)$ are pairwise orthonormal functions on $[0, T]$ and Z_i are pairwise uncorrelated random variables with zero mean and unit variance. If Y is a Gaussian process, then the random coefficients Z_k are independent normal random variables distributed as $Z_k \sim \mathcal{N}(0, 1)$, in which case Eq. 2.44 can be used to sample the process. ϕ_k and λ_k are respectively the eigenfunctions and eigenvalues of the operator T_Y , defined as $T_Y f = \int K_Y(s, t)f(s)ds$. Therefore, as centered Gaussian processes are specified entirely given their covariance functions K_Y , we can see the KKL expansion as a *diagonalisation* of Y .

The solution to Eq. 2.43 for the one-dimensional Brownian bridge process is known. The result, trivially extended to a d -dimensional Brownian bridge process with diffusion constant D , can be written as

$$\mathbf{B}(t) = \mathbf{x}_0 + (\mathbf{x}_T - \mathbf{x}_0) \frac{t}{T} + \sqrt{2D} \sum_{i=1}^{\infty} \mathbf{Z}_k \sqrt{\lambda_k} \phi_k(t) \quad (2.45)$$

where $\mathbf{Z}_k \in \mathbb{R}^d$, $k = 1, \dots, \infty$ are independent d -dimensional normal random variables

distributed as $\mathbf{Z}_k \sim \mathcal{N}(0, \mathbb{1}_d)$, and where

$$\phi_k(t) = \sqrt{\frac{2}{T}} \sin(t/\sqrt{\lambda_k}) \quad (2.46a)$$

$$\lambda_k = \frac{T^2}{\pi^2 k^2}, \quad (2.46b)$$

and where $\phi_k(t)$ satisfy

$$\int_0^T \phi_k(t) \phi_l(t) dt = \delta_{kl} \quad (2.47)$$

for all $k, l = 1, \dots, \infty$. We will write the components of the random coefficients as $Z_{ik} \equiv (\mathbf{Z}_k)_i$, $i = 1, \dots, d$, and each component is then distributed as $Z_{ik} \sim \mathcal{N}(0, 1)$. The first two terms in Eq. 2.45 are a result of the fact that Brownian bridge, as we have defined it, is not centred and has mean $\langle \mathbf{B}(t) \rangle = \mathbf{x}_0 + (\mathbf{x}_T - \mathbf{x}_0) \frac{t}{T}$.

In the numerical algorithm, we will make use of Eq. 2.45 to generate samples of the Brownian bridge process using, by truncating the sum to N terms and sampling the dN independent random variables $Z_{ik} \sim \mathcal{N}(0, 1)$. The law $\mathbb{P}_{\mathbf{B}}$ therefore becomes a dN -dimensional probability measure $\mathbb{P}_{\mathbf{B}}^N$. Furthermore, we note that ϕ_k also forms an orthonormal basis for the TPE. Therefore we will also expand transition paths $\mathbf{X}^N \in E$ in the same truncated KKL basis of the Brownian bridge process as

$$\mathbf{X}^N(t; A) = \mathbf{x}_0 + (\mathbf{x}_T - \mathbf{x}_0) \frac{t}{T} + \sqrt{2D} \sum_{k=1}^N \mathbf{A}_k \sqrt{\lambda_k} \phi_i(t) \quad (2.48)$$

where $A = (\mathbf{A}_1 \ \mathbf{A}_2 \ \dots \ \mathbf{A}_N) \in \mathbb{R}^{d \times N}$ is a matrix where the columns $\mathbf{A}_k \in \mathbb{R}^d$, $k = 1, \dots, N$ are the coefficients of the expansion. We will refer to \mathbf{A}_k as the modes of $\mathbf{X}^N(t)$, A as the mode matrix and N the *mode truncation*. We have thus introduced a parameterisation $\mathbf{X}(t; A)$ of transition paths, which in turn defines the subspace

$$E^N = \{\mathbf{X}(t; A) : A \in \mathbb{R}^{d \times N}\} \subset E \quad (2.49)$$

of the transition path ensemble. The dN -dimensional probability measure $\mathbb{P}_{\mathbf{X}}^N$ over E^N is found by projecting Eq. 2.23 onto E^N , and we have

$$\frac{d\mathbb{P}_{\mathbf{X}}^N}{d\mathbb{P}_{\mathbf{B}}^N}(A) \propto \exp(-\Phi[\mathbf{X}(t; A)]). \quad (2.50)$$

Accordingly, the pCN acceptance probability would then be

$$A_{\text{pCN}}[A, A'] = \min \{1, \exp(\Phi[\mathbf{X}(t; A)] - \Phi[\mathbf{X}'(t; A')])\}. \quad (2.51)$$

for a given $A, A' \in R^{d \times N}$. We note that the integral in Eq. 2.18 of the relative action Φ

must be computed using a quadrature scheme in numerical applications.

Finally, the pCN rule defined on the KKL mode-representation of transition paths can now be stated as

$$A'_{ik} = \sqrt{1 - \kappa^2} A_{ik}^{(n)} + \kappa_{ik} Z_{ik} \quad (2.52)$$

for each $i = 1, \dots, d$ and $k = 1, \dots, N$, where Z_{ik} is sampled from $\mathcal{N}(0, 1)$. In contrast to Eq. 2.42, the pCN update rule on functional space, here we have introduced a tuneable step-size matrix κ_{ik} for each individual mode coefficient. In Sec. 2.3.5 we will discuss how to use the band-structure of the KKL expansion to find a step-size matrices that improve the MCMC correlation time.

There is an analogous expansion to Eq. 2.48, if we were to instead sample $C_{\mathbf{x}_0}^0([0, T])$, the set of stochastic paths starting at \mathbf{x}_0 with open end-point conditions. In that case, the appropriate expansion of stochastic paths would be the KKL basis of the Wiener process. This would in turn induce an N -dimensional subspace of paths, analogous to Eq. 2.49.

2.3.3 Numerical algorithm

We can summarise the pCN sampling algorithm projected onto the truncated Brownian bridge KKL expansion as follows:

1. Choose the mode truncation parameter N .
2. Choose an initial state $A^{(0)} \in \mathbb{R}^{d \times N}$.
3. Generate a proposal state

$$A'_{ik} = \sqrt{1 - \kappa^2} A_{ik}^{(n)} + \kappa_{ik} Z_{ik} \quad (2.53)$$

for each $i = 1, \dots, d$ and $k = 1, \dots, N$, where Z_{ik} is drawn from $\mathcal{N}(0, 1)$ and κ_{ik} is the step-size matrix.

4. Draw a random number U from $\text{Unif}([0, 1])$.
 - If $U < A_{\text{pCN}}[A^{(n)}, A']$, then set $A^{(n+1)} = A'$.
 - Otherwise set $A^{(n+1)} = A^{(n)}$.
5. Repeat step 3.

In practice, if the algorithm was run for M iterations, it is often prudent to discard some portion of the first $M_B < M$ of the samples $\{A^{(n)}\}_{n=1}^M$. This is due to the fact that $A^{(0)}$ is in general not a representative sample from the target distribution $\mathbb{P}_{\mathbf{x}}$. We thus let the MCMC ‘burn in’ for M_B iterations, and discard $\{A^{(n)}\}_{n=1}^{M_B}$. In all numerical experiments in this text, we use a burn-in of $M_B/M = 0.3$.

Recall that in the definition of the acceptance probability Eq. 2.51 we must either use the Onsager-Machlup or the Freidlin-Wentzell form of the relative action Φ . As discussed in the introduction, for the purposes of sampling $\mathbb{P}_{\mathbf{X}}$ there is no difference between the two actions. In all numerical results that follows we have used the Onsager-Machlup action, where we have also verified that the Freidlin-Wentzell action lead to the same result.

The evaluation of the acceptance probability A_{pCN} involves the calculation of the integral Eq. 2.19, which necessitates that we evaluate the expansion Eq. 2.48 of truncated sample paths $\mathbf{X}^N(t)$ and their derivatives $\dot{\mathbf{X}}^N(t)$. Let $X_{ik}^N \equiv X_i^N(\Delta tk)$ for $k = 0, \dots, N+1$, where $\Delta t = \frac{T}{N+1}$. Similarly, let $\dot{X}_{ik}^N \equiv \dot{X}_i^N(\Delta tk)$ for $k = 0, \dots, N+1$. Then Eq. 2.19 can be approximated using a trapezoidal rule. If done naively, evaluating all X_{ik}^N is an operation that scales as $O(N^2)$. However, we can exploit the fact that Eq. 2.48 is a similar to a discrete Fourier transform of the mode coefficients A_{ik} .

Let DST_I and DCT_I be the *Discrete Sine Transform of type I* and the *Discrete Cosine Transform of type I* respectively, as defined in [122, 123]. Let $\mathbf{c} \in \mathbb{R}^{N+2}$ be a vector, and we denote its components as c_k , $k = 0, \dots, N+1$. The result of the Discrete Sine Transform is then a vector $\text{DST}_I(\mathbf{c}) \in \mathbb{R}^{N+2}$, and we denote its components as $\text{DST}_I(\mathbf{c})_k$, $k = 0, \dots, N+1$. Let the equivalent hold for DCT_I . It is then possible to show that

$$X_{ik} = \bar{x}_{ik} + \text{DST}_I(\mathbf{c}^i)_{k-1}, \quad k = 1, \dots, N \quad (2.54a)$$

$$X_{ik} = \bar{x}_{ik}, \quad k = 0, N+1 \quad (2.54b)$$

$$\dot{X}_{ik} = \text{DCT}_I(\mathbf{d}^i)_k, \quad k = 0, \dots, N+1 \quad (2.54c)$$

where $\bar{x}_{ik} = (x_0)_i + ((x_T)_i - (x_0)_i)\frac{k\Delta t}{T}$, and \mathbf{c}^i and \mathbf{d}^i are defined as

$$c_0^i = c_{N+1}^i = 0 \quad (2.55a)$$

$$c_k^i = \sqrt{\frac{T}{2}} \frac{1}{\pi k} A_{i(k-1)}, \quad k = 0, \dots, N-1 \quad (2.55b)$$

$$d_0^i = \frac{1}{T} (\bar{x}_{i(N+1)} - \bar{x}_{i0}), \quad d_{N+1}^i = 0 \quad (2.55c)$$

$$d_k^i = \sqrt{\frac{1}{2T}} A_{ik}, \quad k = 2, \dots, N \quad (2.55d)$$

As the DST_I and DCT_I can be computed efficiently using Fast-Fourier Transforms, Eq. 2.54 enables us to reconstruct transition paths from their coefficients with a time-complexity of $O(N \log N)$.

Fig. 2.1 shows the result of $M = 10^8$ iterations of the pCN algorithm to a benchmark problem, with burn-in $M_B/M = 0.3$. We sample the TPE of a 1-dimensional overdamped Langevin dynamics of a particle moving between the minima of an asymmetric double-well potential. See Appendix A.1 for a figure of the potential and a detailed specification of the

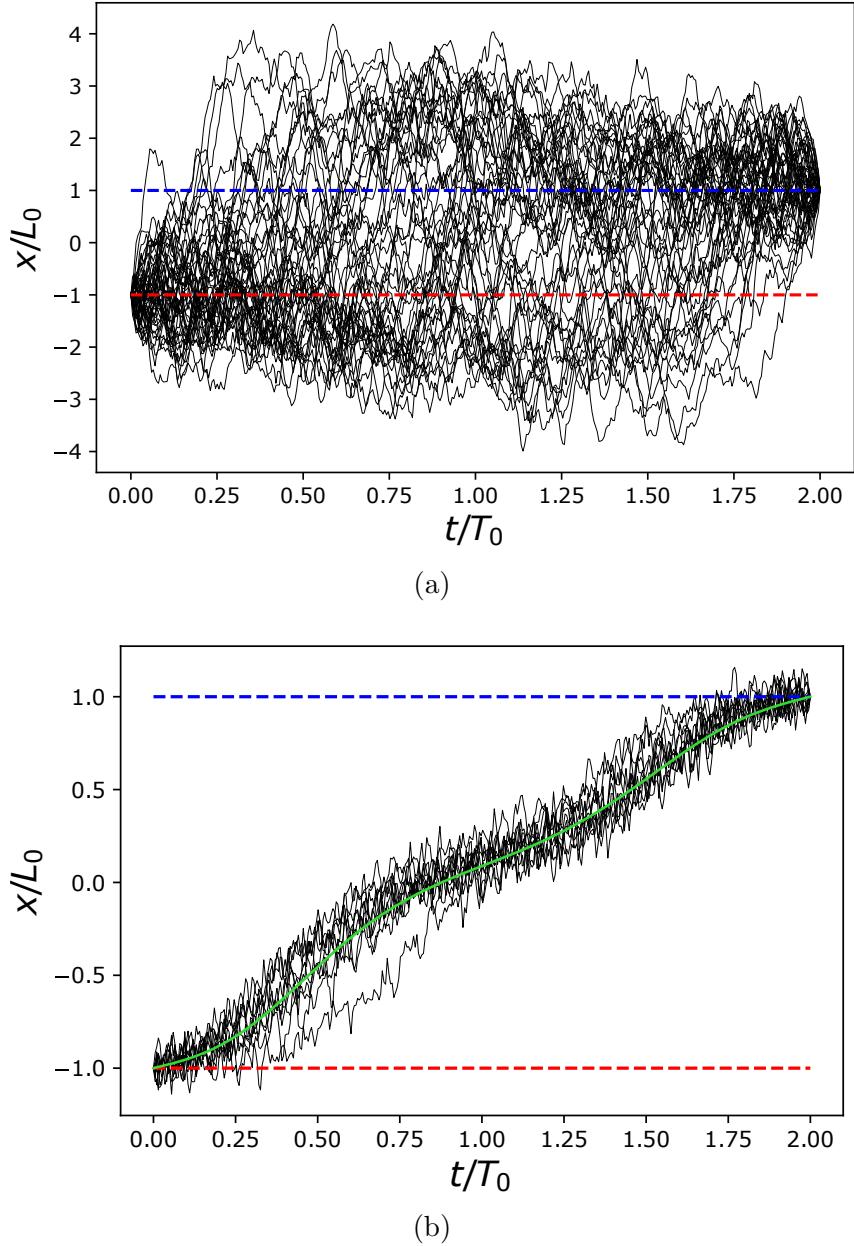


Figure 2.1: Simulation of the 1-dimensional asymmetric double-well system (defined in Appendix A.1) using $M = 10^8$ iterations of the pCN algorithm with burn-in $M_B/M = 0.3$. We sampled transition paths of duration $T/T_0 = 2$ between the minima at $x_0/L_0 = -1$ (red dashed lines) to the minima at $x_0/L_0 = 1$ (blue dashed lines), and with mode truncation $N = 200(T/T_0)$. (a) Shows 50 transition paths at temperature $\theta/\theta_0 = 2$. (b) Shows 10 transition paths at temperature θ/θ_0 , and the instanton (green line).

model. Fig. 2.1a shows a sample of 50 trajectories from the TPE, simulated at temperature $\theta/\theta_0 = 2$, where θ_0 is a reference temperature. The reference temperature θ_0 is defined such that at $\theta = \theta_0$ the thermal fluctuations equal the characteristic energy scale of the confining potential energy (See Appendix A for details on the non-dimensionalisation of the model system). In Fig. 2.1 Fig. 2.1b we show the maximiser of the path-probability density functional Eq. 2.5 (which we call the *instanton*), as well as a sample of 10 trajectories from the TPE simulated at a temperature $\theta/\theta_0 = 0.01$. The duration of the transition paths were $T/T_0 = 2$, where T_0 is a reference time-scale. We computed the instanton by minimising the Onsager-Machlup action of the model system, using a Ritz method as described in Ch. 1. Note that, as mentioned in the introduction, at finite temperatures the instanton must be computed using Onsager-Machlup action, as opposed to the Freidlin-Wentzell action [32, 109]. At temperatures $\theta > \theta_0$, where thermal fluctuations (represented by the noise term in the Langevin equation) dominate over the potential, and we see in Fig. 2.1 (a) that the sample paths transition between the two minima several times. On the other hand, at temperatures $\theta \ll \theta_0$, where we would expect that transition events are rare, the TPE is concentrated around the instanton.

To verify that our implementation of the algorithm correctly sampled the TPE, we verified our results using a standard Euler-Maruyama integrator [80]. Using the latter, we generated samples 10^6 of the TPE by collecting trajectories with end-points within a small interval $\tilde{x}(T) \in [1 - \epsilon, 1 + \epsilon]$ around the right minima, where $\epsilon = 10^{-2}$. We used a temperature θ/θ_0 as in Fig. 2.1a, as it is unfeasible to use the Euler-Maruyama algorithm to sample the TPE for temperature regimes where transition events are rare. We projected the resulting sample paths onto the KKL basis of the Brownian bridge process, and compared the sample with the result of the pCN method. The results are showed in Fig. 2.2, where we plot the marginalised histograms of the mode distributions against each other.

2.3.4 Mode-space band structure

In Fig. 2.2 we see that the marginalised distributions of the modes of stochastic trajectories in the TPE (blue lines) increasingly coincide with normal distributions of unit variance (green lines) at higher mode orders. The latter corresponds to the modes of the Brownian bridge process, as was established in Sec. 2.3.2. Fig. 2.3 shows the absolute normalised covariances $\rho_{kl} = \langle A_k A_l \rangle / \sqrt{\langle A_k^2 \rangle \langle A_l^2 \rangle}$ of the modes. Here we see the modes also rapidly decorrelate, and are numerically negligible beyond the first 30 modes. Fig. 2.2 and Fig. 2.3 are a first indication that modes of transition paths are approximately distributed like a Brownian bridge process for large mode orders k . As we discuss and show below, this holds for *any* Itô diffusion with additive noise, regardless of the drift-field \mathbf{b} .

The law $\mathbb{P}_{\mathbf{X}}$ over the TPE of an Itô diffusion can be seen as a non-linear transformation

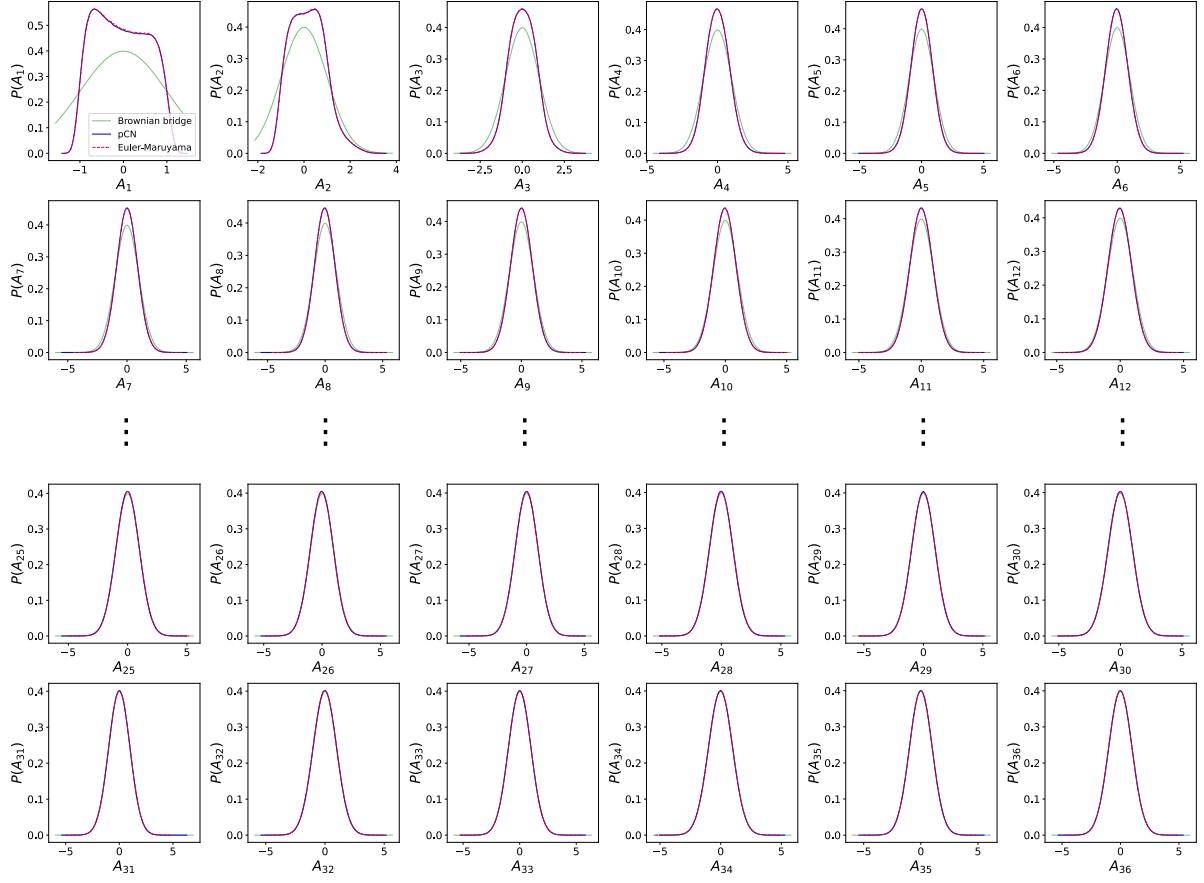


Figure 2.2: Simulation of the 1-dimensional asymmetric double-well system using both the pCN and the Euler-Maruyama methods. The resulting samples of the latter were projected onto the KKL basis of the Brownian bridge process, and the figure shows the marginalised distribution over the coefficients A_k of the sampled TPE. In each sub-plot, the green line are the distributions of a centred normal with unit-variance $Z_k \sim \mathcal{N}(0, 1)$, which corresponds to the modes of the Brownian bridge process. The parameters of the simulation are the same as those of Fig. 2.1 (a).

of \mathbb{P}_B , the law of the Brownian bridge process. They are related via Eq. 2.23 in terms of the density Eq. 2.16, where the latter is an expression entirely in terms of the drift-field. Therefore, the difference in statistical ‘information’ contained in \mathbb{P}_X relative to \mathbb{P}_B is exactly the drift-field \mathbf{b} .

Furthermore, the difference in distribution between \mathbb{P}_X and \mathbb{P}_B is located in the lower-end of the mode-spectrum. For the 1-dimensional benchmark problem, in Fig. 2.3 we can see a division in mode space between a low-frequency and high-frequency band, where the division between the two bands can be conservatively assigned to around $k_l = 40$. The low-frequency band thus correspond to the first k_l modes in the expansion Eq. 2.48, and the high-frequency band correspond to all modes $k > k_l$. We can consider the low-frequency modes the ‘smooth’ and the high-frequency modes the ‘rough’ parts of a stochastic trajectory, as visualised in Fig. 2.4. The high-frequency noise is also centred, such that it does not effect the overall trajectory. The underlying shape of the path is

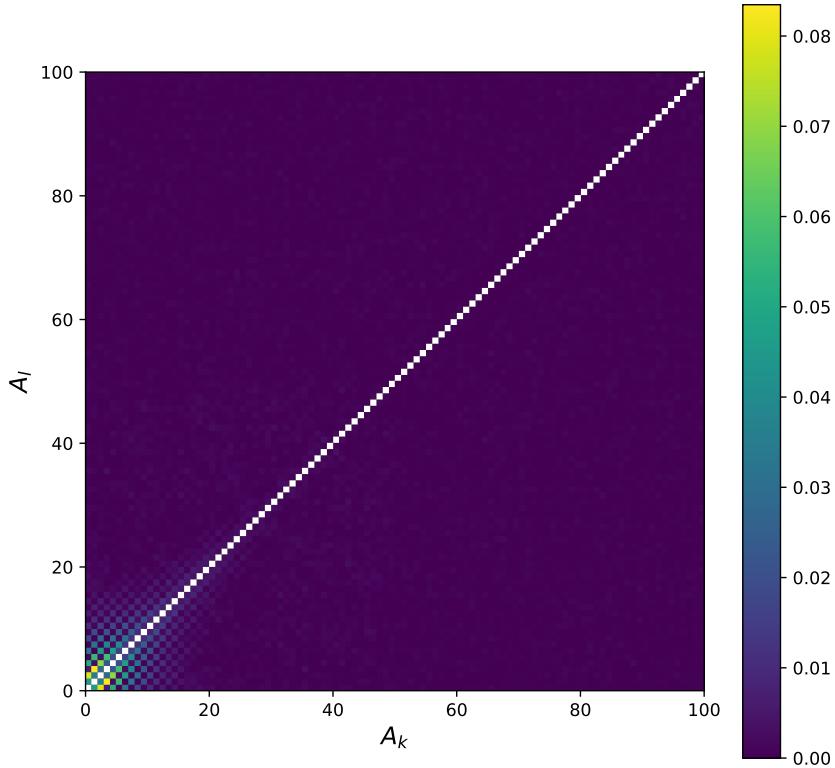


Figure 2.3: The absolute normalised covariances $\rho_{kl} = \langle A_k A_l \rangle / \sqrt{\langle A_k^2 \rangle \langle A_l^2 \rangle}$ of the first 100 modes of the transition paths of the 1-dimensional asymmetric double-well system. As $\rho_{kk} = 1$, $k = 1, \dots, N$ by definition, we have excluded the diagonal from the plot. We sampled the TPE using the pCN algorithm, with parameters $T/T_0 = 2$, $\theta/\theta_0 = 2$ and $N = 200(T/T_0)$ with $M = 10^8$ iterations. We see that there is a band separation between the low and high mode-numbers, where modes beyond the the $k \in [0, 40]$ quadrant are largely uncorrelated.

thus only encoded in the low-frequency modes. In the dynamics of an Itô diffusion Eq. 2.3, the stochastic term $\sigma d\mathbf{W}$ dominates over deterministic drift-field \mathbf{b} at short-time scales. In other words, \mathbf{b} only affects the system on long time-scales. Therefore we would intuit that stochastic realisations \mathbf{X} , of the Itô diffusion equation, look like they are undergoing drift-less free diffusion as we ‘zoom in’ on the trajectory. Therefore, if \mathbf{b} is smooth, we would expect that $\mathbb{P}_{\mathbf{X}}$ and $\mathbb{P}_{\mathbf{B}}$ only behave differently in the ‘smooth’ low-frequency band, and behave identically in ‘rough’ high-frequency band. Further numerical evidence of the band separation is provided indirectly in Sec. 2.3.5 and in later sections in Fig. 2.10 and Fig. 2.9, in addition to the results already discussed for the 1-dimensional asymmetric double-well system above.

The band separation of the sampling method serves as an explanation for why the pCN method does not suffer from a curse of dimensionality as $N \rightarrow \infty$. As modes k beyond some cutoff k_l are approximately independently distributed normal random variables, the effective dimension of the target sampling space is dk_l , regardless of any choice of mode

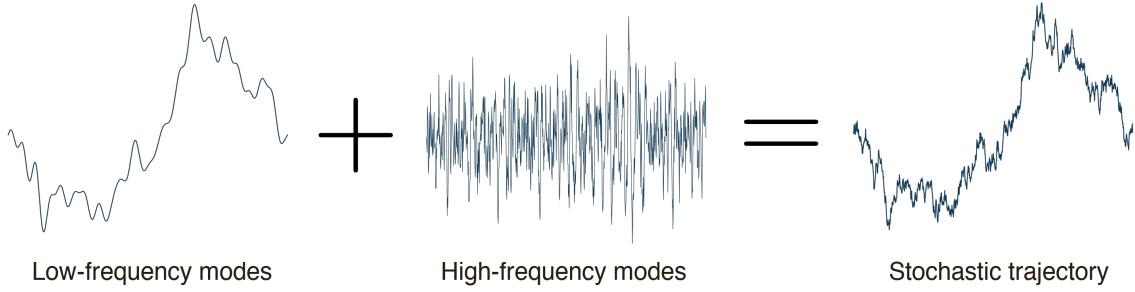


Figure 2.4: Illustration of the spectral decomposition of stochastic transition paths, in the Kosambi-Karhunen-Loeve basis of the Brownian bridge process. Stochastic paths can be seen as the superposition of low-frequency modes, which resemble smooth trajectories, and zero-mean high-frequency noise. The general shape of the path is thus encapsulated in the former, whilst the latter adds roughness to the path. These same arguments can be made for a stochastic trajectory open end-point conditions, where the underlying Gaussian reference measure is then the Wiener process.

truncations $N > k_1$.

Statements analogous to the above would apply if we instead considered the spectrum of the law of an unpinned Itô diffusion, with sample space $C_{\mathbf{x}_0}^0([0, T])$, the set of stochastic paths starting at \mathbf{x}_0 with open end-point conditions at $t = T$. In that case, using identical arguments, we would find that the spectrum of the $\mathbb{P}_{\mathbf{x}}$ would resemble $\mathbb{P}_{\mathbf{W}}$ in the high-frequency band, where the latter is the law of the Wiener process. Below we will provide an analytical argument for the case of an unpinned Itô diffusion.

In the following we expand a unpinned Langevin processes of the form Eq. 2.3 in the KKL basis of the Wiener process \mathbf{W} . For simplicity we only consider the 1-dimensional case, but the results below readily generalises to higher dimensional processes. The KKL expansion of the 1-dimensional Wiener process is

$$W(t) = \mathbf{x}_0 + \sqrt{2D} \sum_{k=1}^{\infty} Z_k \sqrt{\nu_k} \chi_k(t) \quad (2.56)$$

where Z_k , $k = 1, \dots, \infty$ are independent d -dimensional normal random variables distributed as $\mathbf{Z}_k \sim \mathcal{N}(0, \mathbb{I}_d)$, and where

$$\chi_k(t) = \sqrt{\frac{2}{T}} \sin(t/\sqrt{\nu_k}) \quad (2.57a)$$

$$\nu_k = \frac{T^2}{\pi^2 \left(k - \frac{1}{2}\right)^2}, \quad (2.57b)$$

and where $\chi_k(t)$ satisfy $\langle \chi_k, \chi_l \rangle = \delta_{kl}$, where we have defined

$$\langle f, g \rangle = \int_0^T f(t)g(t)dt. \quad (2.58)$$

for any two functions $f, g : [0, T] \rightarrow \mathbb{R}$. We begin with the integral form of the Itô diffusion equation Eq. 2.4

$$X(t) = \int_0^t b(X(t'))dt' + \sqrt{2D} \int_0^t dW(t') \quad (2.59)$$

where $b : \mathbb{R} \rightarrow \mathbb{R}$ is the drift-field, $t \in [0, T]$ and the second term of Eq. 2.59 is an Itô integral. We assume, without loss of generality, that $X(0) = 0$. We now expand the Wiener process in its KKL expansion, and truncate the sum up to order N to approximate Eq. 2.59 as

$$X^N(t) = \int_0^t b(X^N(t'))dt' + \int_0^t dW^N(t') \quad (2.60)$$

where W^N is the truncation of Eq. 2.56, X^N is the truncated stochastic process, and where the second term in Eq. 2.60 is now a Riemann-Stieltjes integral. It is a theorem (known as the *Wong-Zakai theorem*) [124, 125, 126] that as $N \rightarrow \infty$ the sequence of solutions to Eq. 2.60 converges uniformly in probability to the solution of Eq. 2.59².

We will expand X^N as

$$X^N(t) = \sqrt{2D} \sum_{k=1}^N A_k \sqrt{\nu_k} \chi_k(t). \quad (2.61)$$

As $X^N(t)$ and $W^N(t)$ are smooth for finite N , we can take the derivative of Eq. 2.60 to find

$$\dot{X}^N(t) = b(X^N(t)) + \dot{W}^N(t). \quad (2.62)$$

We can expand the $\dot{X}^N(t)$ and $\dot{W}^N(t)$ as

$$\dot{X}^N(t) = \sqrt{2D} \sum_{k=1}^N A_k \psi_k(t) \quad (2.63a)$$

$$\dot{W}^N(t) = \sqrt{2D} \sum_{k=1}^N Z_k \psi_k(t) \quad (2.63b)$$

where

$$\psi_k(t) = \sqrt{\nu_k} \dot{\chi}_k(t) = \sqrt{\frac{2}{T}} \cos(t/\sqrt{\nu_k}). \quad (2.64)$$

As $\langle \psi_k, \psi_l \rangle = \delta_{kl}$, we have that ψ_k forms an orthonormal basis over the space of paths. We

²Note that the Wong-Zakai theorem holds for the Stratonovich formulation of Eq. 2.59. However, for additive noise the Itô and Stratonovich formulations are equivalent.

can therefore expand the drift-field as

$$b(X^N(t)) = \sqrt{2D} \sum_{k=1}^N b_k \psi_k(t) \quad (2.65)$$

where

$$b_k = \frac{1}{\sqrt{2D}} \langle \psi_k, b(X^N) \rangle = \int_0^T \psi_k(t) b(X^N(t)) dt. \quad (2.66)$$

From Eq. 2.62, and the orthonormality of ψ_k , we get

$$A_k = b_k + Z_k, \quad (2.67)$$

where $k = 1, \dots, N$. This is a spectral version of the Itô diffusion equation, where we have projected Eq. 2.59 onto the truncated KKL spectrum of the Wiener process. Here the mode coefficients A_k are random variables expressed in terms of the drift modes b_k and the modes of the Wiener process Z_k . From Eq. 2.66 we see that $b_k = b_k(A)$ is in general a non-linear equation in terms of the mode coefficients $A = (A_1, \dots, A_N)$. We note that Eq. 2.67 can in principle be used to sample the Itô diffusion equation, by drawing N random numbers $Z_k \sim \mathcal{N}(0, 1)$ and then solving Eq. 2.67 for A .

Any finite choice of the number of modes N will lead to truncation errors, but for increasing N Eq. 2.67 will probabilistically converge to the true infinite-dimensional distribution. For a given X^N , b_k can be seen as the Fourier coefficients of b . Now, using the known convergence properties of Fourier expansions, we have that

$$|b_k| \leq \frac{C(D, T)}{k} \quad (2.68)$$

where $C(D, T)$ is some function of the system parameters. As Z_k is always of order 1, Eq. 2.67 tells us that

$$A_k \approx Z_k \quad (2.69)$$

for $k > k_{\parallel}$ for some value of k_{\parallel} , and sufficiently large $N > k_{\parallel}$. We thus see that for a given system, there are always only a finite number of physically relevant modes distinct from free diffusion dynamics. Therefore, the effective dimensionality of the target sample space is dk_{\parallel} , for a general d -dimensional Itô diffusion. In general k_{\parallel} be dependent on the drift-field b , and from Eq. 2.66 we see that it will be an increasing function of the duration T and the diffusivity D . This latter fact is intuitive, as the influence of the drift-field in the dynamics increasingly dominates over the noise as the noise-amplitude $\sqrt{2D}$ is lowered and T is increased.

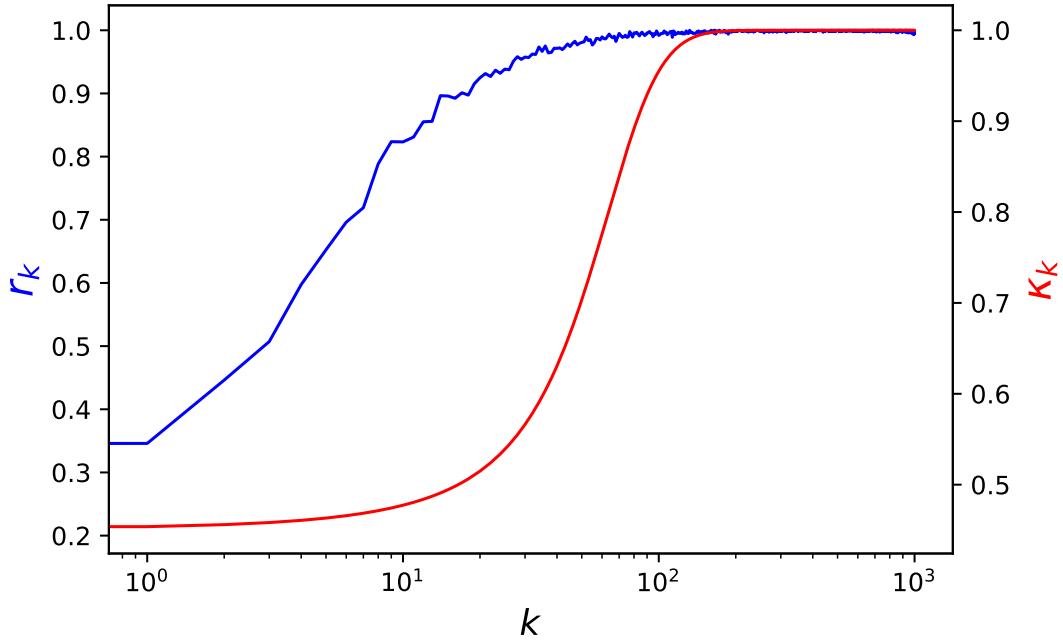


Figure 2.5: (Red line) The modal acceptance rates r_k^{acc} for the 1-dimensional asymmetric double-well system, with system parameters identical to those in Fig. 2.1a. (Blue line) A sigmoidal step-size profile κ_k .

2.3.5 Mode-dependent step-sizes

The band-structure of the mode-spectrum of Itô processes can potentially be exploited to improve the sampling algorithm. Here, as a simple first attempt, we will use the band-structure to adapt the step-sizes in the pCN algorithm for each individual mode. To construct an adapted step-size matrix κ_{ik} , we need to first have a method by which to estimate the extent to which each mode is distributed like a Gaussian. Note that if $\kappa_{ik} = 1$ for a given i, k , then the proposal Eq. 2.53 at each iteration is an independent random sample from $\mathcal{N}(0, 1)$.

Consider a modified pCN algorithm that, at each iteration step n , only updates a single mode A_{ik} at a time. The algorithm updates the mode in sequence, such that after dN iterations all of the modes A_{ik} , $i = 1, \dots, d$, $k = 1, \dots, N$ will have been updated once. Let r_{ik} be the *modal acceptance rate* of the iterations of the mode A_{ik} , which is the ratio of accepted proposals to the number of iterations $M/(dN)$, where M is the number of iterations of the whole algorithm. Similarly, let r be the acceptance rate of all modes in the regular pCN algorithm. Figure 2.5 plots the modal acceptance rates (blue line) for the 1-dimensional asymmetric double-well system, with system parameters identical to those in Fig. 2.1a, mode truncation $N = 1000$ and iterations $M = 1000N$. We see that

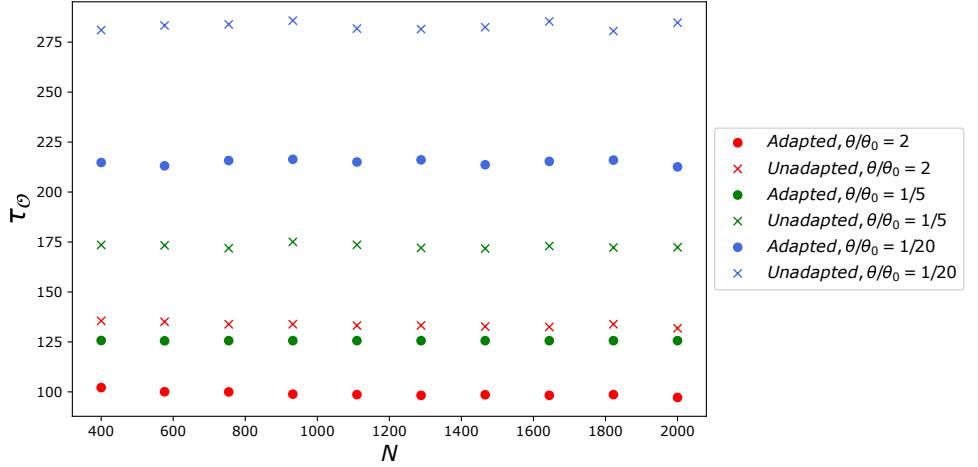


Figure 2.6: The auto-correlation times $\tau_{\mathcal{O}}$ of MCMC runs of the asymmetric double-well system at temperatures $\theta/\theta_0 = 2, 1/5, 1/20$ with adapted and unadapted modal step-sizes κ_i , for varying mode truncations N .

$r_k \approx 1$ for modes $k \geq 50$. Therefore we may consider a sigmoidal step-size profile

$$\kappa_k = \kappa_{\min} + \frac{\kappa_{\max} - \kappa_{\min}}{1 + \exp(-\sigma_{\kappa}(k - k_{|}))} \quad (2.70)$$

with $\kappa_{\min} = 0.4$, $\kappa_{\max} = 1$, $k_{|} = 50$ and $\sigma_{\kappa} = 0.05$, which is plotted in Fig. 2.5 (blue line). The base-line step-size κ_{\min} was computed by finding an approximate constant $\kappa_k = \bar{\kappa}$ that yielded an acceptance rate $r \approx 0.3$ in the regular pCN algorithm.

We benchmarked MCMC runs with adapted step-size profiles against runs with constant steps-sizes by considering their relative *auto-correlation times*. Let $\mathcal{O}[\mathbf{X}]$ be some observable, and $\mathbf{X}^{(n)}$, $n = 1, \dots, M$ the result of an MCMC run. Then the expectation $\langle \mathcal{O}[\mathbf{X}] \rangle$ can be estimated using Eq. 2.26, and the variance as

$$\text{Var}(\mathcal{O}[\mathbf{X}]) = \langle (\mathcal{O}[\mathbf{X}] - \langle \mathcal{O}[\mathbf{X}] \rangle)^2 \rangle = \frac{\tau_{\mathcal{O}}}{M} \sum_{n=1}^M (\mathcal{O}[\mathbf{X}^{(n)}] - \langle \mathcal{O}[\mathbf{X}] \rangle)^2, \quad (2.71)$$

where $\tau_{\mathcal{O}}$ is the *integrated autocorrelation time* for the chain $\mathcal{O}[\mathbf{X}^{(n)}]$, $n = 1, \dots, M$. It carries the interpretation of the amount of iterations required for the MCMC to ‘forget’ where it began, such that the effective number of samples is $M/\tau_{\mathcal{O}}$. See [127, 128] for detailed definitions and how to estimate $\tau_{\mathcal{O}}$. In Fig. 2.6 we show the results of the benchmark runs. For mode truncations varying between $N = 400$ and $N = 2000$, we computed the auto-correlation times for MCMC runs with an adapted step-size κ_k as in Eq. 2.70, as well as an unadapted step-size $\kappa_k = \kappa_{\min}$. We ran simulations at temperatures $\theta/\theta_0 = 2, 1/5, 1/20$, with $\kappa_{\min} = 0.4, 0.35, 0.3$ respectively, and $\kappa_{\max} = 1$, $k_{|} = 50$ and $\sigma_{\kappa} = 0.05$ for all temperatures. We see that the adapted step-size runs enjoy lower autocorrelation times.

The above is a modest attempt at utilising the band-structure to improve autocorrelation times. A more promising avenue would be to apply a multi-level MCMC algorithm [129, 130, 131, 132], which consists of a series of M_{multi} chains with a hierarchy of mode truncations $N_i, i = 1, \dots, M_{\text{multi}}$ such that $N_i < N_{i+1}$. The MCMC can be efficiently run on the lowest-dimensional chain, and then be *upsampled* via the multi-level scheme, such that we generate samples of the highest $N_{M_{\text{multi}}}$ -dimensional system.

2.3.6 Generalised end-point conditions

So far we have considered transition path ensembles with fixed end-point conditions. We may consider a TPE of a more general form

$$E_{R_0}^{R_T}([0, T]) = \{\mathbf{X} : \mathbf{X} \in C_{\mathbf{x}_0}([0, T]) \text{ and } \mathbf{X}(T) = \mathbf{x}_T \text{ where } \mathbf{x}_0 \in R_0, \mathbf{x}_T \in R_T\}. \quad (2.72)$$

In other words, $E_{R_0}^{R_T}([0, T])$ is the set of stochastic paths that start in $R_0 \subset \mathbb{R}^d$ and end in $R_T \subset \mathbb{R}^d$. In applications, R_0 and R_T may, for example, be chosen to lie within the basins of attraction of fixed points of the drift-field \mathbf{b} .

A probability measure over $E_{R_0}^{R_T}([0, T])$ can be constructed by simply grafting a probability distribution over the initial condition onto the path-probability measure $\mathbb{P}_{\mathbf{X}}$. Let π_0 be a probability measure over R_0 , with density $p_0 \sim \pi_0$. Let $\tilde{P}_{\mathbf{X}}$ be a probability measure over $E_{R_0}^{R_T}([0, T])$, defined in terms of the density $\tilde{P}_{\mathbf{X}} \sim \tilde{\mathbb{P}}$

$$\tilde{P}_{\mathbf{X}}[\mathbf{x}] \propto p_0(\mathbf{x}(0)) P_{\mathbf{X}}[\mathbf{x}] \quad (2.73)$$

where we have restricted the domain of $\tilde{P}_{\mathbf{X}}[\mathbf{x}]$ to paths that satisfy $\mathbf{x}(0) \in R_0$ and $\mathbf{x}(T) \in R_T$.

We can modify the pCN algorithm to sample $\tilde{P}_{\mathbf{X}}$. In addition to the coefficient matrix $A = (\mathbf{a}_1 \ \mathbf{a}_2 \dots \mathbf{a}_N)$, the MCMC state will now also be specified by the starting- and end-points of the transition path \mathbf{x}_0 and \mathbf{x}_T respectively. We define the new update rule

$$\mathbf{a}'_i = \sqrt{1 - \kappa^2} \mathbf{a}^{(n)} + \kappa \mathbf{Z}_i \quad (2.74a)$$

$$\text{Draw } \mathbf{x}'_0 \text{ from } \pi_0 \quad (2.74b)$$

$$\text{Draw } \mathbf{x}'_T \text{ from } \pi_T \quad (2.74c)$$

where π_T is a distribution over R_T that must be specified. Let $p_T \sim \pi_T$, then the transition kernel density is

$$\tilde{Q}[\mathbf{X} \rightarrow \mathbf{Y}] = p_0(\mathbf{Y}(0)) p_T(\mathbf{Y}(T)) Q_{\text{pCN}}[\mathbf{X} \rightarrow \mathbf{Y}]. \quad (2.75)$$

From Eq. 2.31 we find the acceptance probability

$$\tilde{A}[\mathbf{X}, \mathbf{Y}] = \min \left\{ 1, e^{\Phi[\mathbf{X}] - \Phi[\mathbf{Y}]} \frac{p_T(\mathbf{X}(T))}{p_T(\mathbf{Y}(T))} \right\}, \quad (2.76)$$

where transition paths are expanded as

$$\mathbf{X}^N(t; A, \mathbf{x}_0, \mathbf{x}_T) = \mathbf{x}_0 + (\mathbf{x}_T - \mathbf{x}_0) \frac{t}{T} + \sqrt{2D} \sum_{i=1}^N \mathbf{a}_i \sqrt{\lambda_i} \phi_i(t). \quad (2.77)$$

The update rule Eq. 2.74 has not been numerically tested, but is rather offered here as a proof-of-concept to show that it is possible to implement generalised initial- and final end-point configurations using the pCN method. Equation 2.74b and Eq. 2.74c are here independent samplers, and in practice implementing random-walk procedures in R_0 and R_T might lead to better convergence.

Finally, some applications may require that we let r of the d degrees of freedom have open end-point conditions (In other words, $R_T = \mathbb{R}^d$). Let X_i be the components of the stochastic process \mathbf{X} , and we consider transition paths where we let $X_i, i = 1, \dots, r$ have open end-point configurations. We may then adapt the pCN algorithm of Sec. 2.3.3 by expanding $X_j, j = r+1, \dots, d$ in the KKL basis of the Brownian bridge process as before, and expanding $X_i, i = 1, \dots, r$ in the KKL basis of the *Wiener process*.

2.4 Sampling multi-modal transition path ensembles

Many applications of MCMC methods involve state-spaces that consist of large or very complicated probability landscapes. In particular, the target probability measure may be concentrated around several local maxima that are far-removed from each other in state-space. In such cases, one of the major difficulties that MCMC methods face is either the slow or complete failure of the chain to mix between the local maxima of the target measure. Inherently, the issue stems from the fact that the samples generated by the MCMC are not independent, but correlated. That is, a future state $X^{(n+1)}$ will often lie within the general vicinity of the current state $X^{(n)}$ in the chain. If $X^{(n)}$ is in a region of locally high probability in state-space, then the likelihood of the chain to traverse to another local maximum is low if the two regions are separated by regions of low probability. In this section, we will develop an algorithm which allows for simultaneously sampling multiple local maxima. We will begin by first defining the algorithm for finite dimensional spaces, and then in subsequent subsections we will develop the necessary theory to define the method on the transition path ensembles of Itô diffusions.

As a motivating example, consider a target measure with a 2-dimensional state space with two local maxima. If the support of the measure is heavily concentrated around the

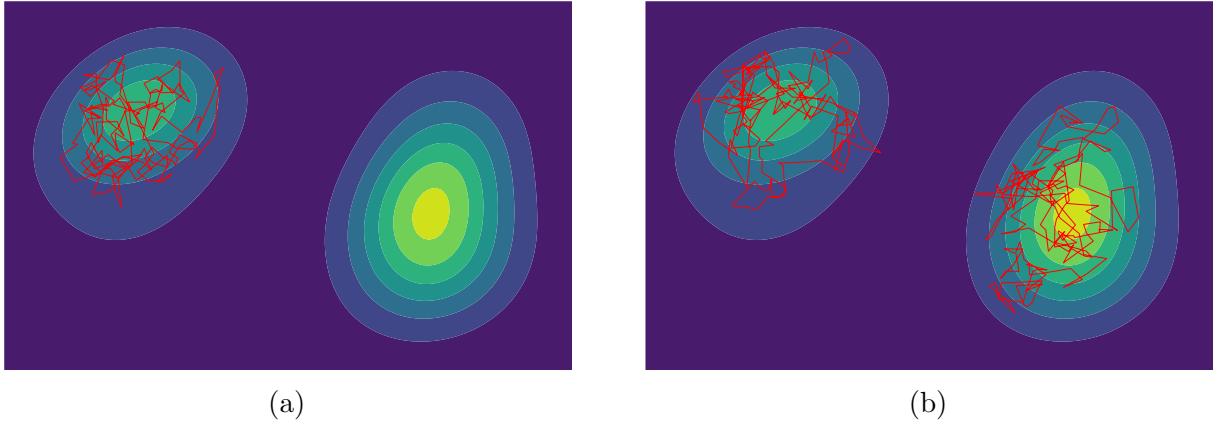


Figure 2.7: MCMC samples of a 2-dimensional distribution with two segregated regions of high probability. (a) An illustration of an RMW scheme that fails to mix between the two regions. (b) An illustration of the TMC scheme, that teleports between the two regions to sample both simultaneously.

two maxima, and if the two regions are sufficiently segregated from each other, then we expect slow mixing to be likely during sampling. Fig. 2.7a illustrates an example of an RMW sampling such a distribution. Whilst the chain will thoroughly sample the locally connected region, it might only rarely traverse between the two regions. The resulting sample will not be representative of the target measure, as it is incorrectly skewed towards one maximum at the expense of the other.

The issue of slow mixing can be mitigated via a simple modification of the MCMC proposal rule. Consider a general target measure \mathbb{P} with density $P \sim \mathbb{P}$ over \mathbb{R}^d , and an MCMC scheme defined by the transition kernel $T(\mathbf{x} \rightarrow \mathbf{y})$, which has \mathbb{P} as its invariant measure. Furthermore, let $p_T \in [0, 1]$ and let $T_T(\mathbf{x} \rightarrow \mathbf{y})$ be an additional MCMC scheme with the same target measure we will define below. We define the superposition of the two schemes as

$$T_{\text{TMC}}(\mathbf{x} \rightarrow \mathbf{y}) = (1 - p_{\text{teleport}})T(\mathbf{x} \rightarrow \mathbf{y}) + p_T T_T(\mathbf{x} \rightarrow \mathbf{y}). \quad (2.78)$$

Equation 2.78 defines a Markov chain with \mathbb{P} as its invariant measure, that, with probability p_{teleport} transitions using $T_T(\mathbf{x} \rightarrow \mathbf{y})$, and with probability $1 - p_{\text{teleport}}$ transitions using $T(\mathbf{x} \rightarrow \mathbf{y})$. The key idea behind the algorithm is to fashion $T_T(\mathbf{x} \rightarrow \mathbf{y})$ such that the MCMC can ‘teleport’ between segregated local regions of high probability, will therefore call $T_T(\mathbf{x} \rightarrow \mathbf{y})$ the *teleportation kernel*. Let \mathbb{P}_T be a probability measure we call the *teleportation measure*, with density $P_T \sim \mathbb{P}_T$, that we can efficiently draw independent samples from. We now construct $T_T(\mathbf{x} \rightarrow \mathbf{y})$ using the Metropolis-Hastings scheme Eq. 2.29, with proposal kernel

$$Q_T(\mathbf{x} \rightarrow \mathbf{y}) = P_T(\mathbf{y}), \quad (2.79)$$

and from Eq. 2.31 we find the acceptance kernel

$$A_T(\mathbf{x}, \mathbf{y}) = \min \left\{ 1, \frac{P[\mathbf{y}] P_T(\mathbf{x})}{P[\mathbf{x}] P_T(\mathbf{y})} \right\} \quad (2.80)$$

The second term of Eq. 2.78 thus draws proposal states independently from \mathbb{P}_T . Now if \mathbb{P}_T is some suitable choice of teleportation measure that concentrates around the local maxima of the target measure \mathbb{P} , the resulting MCMC scheme Eq. 2.78 will be able to ‘teleport’ between maxima with some probability at each iteration. We call this the *teleporter MCMC* scheme (TMC). Similar algorithms have been proposed in [95, 96]. Fig. 2.7b shows an illustrative example of a TMC run.

Note that the TMC requires knowledge of the extrema of the distribution. However, as numerical optimisation is in general computationally cheaper than MCMC sampling, finding the maxima of the target measure is often feasible. In the case of paths-spaces, Ritz methods of the kind described in Ch. 1 can be used. Once the maxima have been found, there are in principle many choices for \mathbb{P}_T . However, from Eq. 2.80 we see that care must be taken so that the teleportation measure does not generate proposals that are probabilistically unfavourable.

One way to ensure the suitability of the teleportation measure is by constructing it as a superposition of multi-variate Gaussians, that each locally approximates \mathbb{P} around each maximum. Let $V(\mathbf{x}) = \log P(\mathbf{x})$ be the exponential scaling of the target measure density, and let $\mathbf{x}^{[\alpha]}$, $\alpha = 1, \dots, K$ be the K local maxima of $P(\mathbf{x})$. Then we can approximate $P(\mathbf{x})$ around $\mathbf{x}^{[\alpha]}$ by Taylor expanding $V(\mathbf{x})$ to second order. We get

$$P_G^{[\alpha]}(\mathbf{x}) = \frac{1}{Z^{[\alpha]}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mathbf{x}^{[\alpha]})^T H^{[\alpha]} (\mathbf{x} - \mathbf{x}^{[\alpha]}) \right) \quad (2.81)$$

where $Z^{[\alpha]}$ is a normalisation constant and

$$H_{ij}^{[\alpha]} = (\partial_i \partial_j V)|_{\mathbf{x}=\mathbf{x}^{[\alpha]}}. \quad (2.82)$$

The Hessian matrix $H^{[\alpha]}$ is, in this context, known as the *precision matrix* of the multi-variate Gaussian $\mathbb{P}_G^{[\alpha]}$. Correspondingly, we have that $(H^{[\alpha]})^{-1}$ is the covariance matrix, such that we can write $\mathbb{P}_G^{[\alpha]} \sim \mathcal{N}(\mathbf{x}^{[\alpha]}, (H^{[\alpha]})^{-1})$. Using Eq. 2.81 we construct the teleportation measure as

$$P_T(\mathbf{x}) = \sum_{\alpha=1}^K w_{\alpha} P_G^{[\alpha]}(\mathbf{x}) \quad (2.83)$$

where w_{α} are tuneable weights that must satisfy $\sum_{\alpha=1}^K w_{\alpha} = 1$.

In the following subsections, we will be generalising the TMC to infinite-dimensional path-spaces. In this setting, our aim is to effectively sample the TPEs of systems that

have multiple competing transition channels. At low diffusivity (low temperature), these channels concentrate around the *local instantons* $\mathbf{x}^{[\alpha]}$ of the Onsager-Machlup action. The majority of the treatment below will revolve around generalising Eq. 2.83.

2.4.1 Second-order variational expansions of stochastic action functionals

In the finite-dimensional case, we locally approximated the target measure by Taylor expanding its logarithm to second order around a local maximum, which resulted in a Gaussian distribution $\mathcal{N}(\mathbf{x}^{[\alpha]}, (Q^{[\alpha]})^{-1})$ centred around the maximum. Here, we apply the analogous procedure to approximate $\mathbb{P}_{\mathbf{X}}$, the path-probability measure over the transition path ensemble. This will result in a Gaussian distribution on path-space, i.e. Gaussian processes. Analogously to the finite-dimensional case, Gaussian processes can be expressed as $\mathcal{N}(\bar{\mathbf{x}}, \mathcal{H}^{-1})$, where the path $\bar{\mathbf{x}}$ is the mean of the process, \mathcal{H} is a self-adjoint linear operator known as the *precision operator*, and \mathcal{H}^{-1} is the *covariance operator*. The density of such a Gaussian process, with respect to a fictitious infinite-dimensional Lebesgue measure, can then be written as

$$P[\mathbf{x}] \propto \exp\left(-\frac{1}{2}\langle \mathbf{x} - \bar{\mathbf{x}}, \mathcal{H}(\mathbf{x} - \bar{\mathbf{x}}) \rangle\right) \quad (2.84)$$

where $\langle \mathbf{f}, \mathbf{g} \rangle = \sum_i \int_0^T f_i(t)g_i(t)dt$. Note that $(\mathbf{x} - \bar{\mathbf{x}})$ is not an argument of \mathcal{H} in Eq. 2.84). Rather \mathcal{H} is acting on $(\mathbf{x} - \bar{\mathbf{x}})$ as a linear operator. We will thus approximate path-probabilities as Gaussian processes using a variational approach. From the path-integral perspective, this is often known as a semi-classical approximation [89, 90, 91, 92, 93, 94]. A non-variational approach was previously shown in [36], which approximates non-Gaussian processes by minimising a Kullback-Leibler divergence between the target process and the approximating Gaussian.

Recall that we expressed the path-probability density Eq. 2.6 of an Itô diffusion in terms of its exponential scaling, which had either the Freidlin-Wentzell and Onsager-Machlup forms. For what follows, either of the stochastic actions can be used³, for the same reasons discussed in the introduction. However, it is important to note that all local instantons must be computed using the OM action.

We write the action in terms its integrand as

$$S[\mathbf{x}(t)] = \int_0^T L(\mathbf{x}(t), \dot{\mathbf{x}}(t))dt \quad (2.85a)$$

$$L(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2D}|\dot{\mathbf{x}} - \mathbf{a}|^2 + \frac{1}{2}\nabla \cdot \mathbf{a}. \quad (2.85b)$$

³In all numerical results presented in this text we used the OM action.

By Taylor-expanding the integrand Eq. 2.85b, we can in turn construct a second-order expansion of the action. Firstly, we introduce the notation that $\frac{\partial L}{\partial \mathbf{x}}(\bar{\mathbf{x}}, \dot{\bar{\mathbf{x}}})$ is the vector-function with components $(\frac{\partial L}{\partial \mathbf{x}})_i = \frac{\partial L}{\partial x_i}$ and $\frac{\partial^2 L}{\partial \mathbf{x} \partial \mathbf{x}}(\bar{\mathbf{x}}, \dot{\bar{\mathbf{x}}})$ is the matrix-function with components $(\frac{\partial^2 L}{\partial \mathbf{x} \partial \mathbf{x}})_{ij} = \frac{\partial^2 L}{\partial x_i \partial x_j}$, where $\bar{\mathbf{x}}$ is a given reference path. We will henceforth suppress the arguments of Eq. 2.86. We can expand the action as

$$S[\bar{\mathbf{x}} + \delta \mathbf{x}] = S[\bar{\mathbf{x}}] + J[\delta \mathbf{x}] + \frac{1}{2} H[\delta \mathbf{x}] + O(\delta \mathbf{x}^3). \quad (2.86)$$

where

$$J[\delta \mathbf{x}] = \int_0^T \left\{ \frac{\partial L}{\partial \mathbf{x}} \delta \mathbf{x} + \frac{\partial L}{\partial \dot{\mathbf{x}}} \delta \dot{\mathbf{x}} \right\} dt \quad (2.87a)$$

$$H[\delta \mathbf{x}] = \int_0^T \left\{ \delta \mathbf{x}^T \frac{\partial^2 L}{\partial \mathbf{x} \partial \mathbf{x}} \delta \mathbf{x} + 2 \delta \mathbf{x}^T \frac{\partial^2 L}{\partial \mathbf{x} \partial \dot{\mathbf{x}}} \delta \dot{\mathbf{x}} + \delta \dot{\mathbf{x}}^T \frac{\partial^2 L}{\partial \dot{\mathbf{x}}^T \partial \dot{\mathbf{x}}} \delta \dot{\mathbf{x}} \right\} dt \quad (2.87b)$$

Equation 2.86 is a variational expansion of the action around the reference path $\bar{\mathbf{x}}(t)$, expressed in terms of a variational test-function $\delta \mathbf{x}$. We will now recast Eq. 2.86 in terms of self-adjoint operators using integration-by-parts and $\delta \mathbf{x}(0) = \delta \mathbf{x}(T) = 0$. We also note that $\langle \mathbf{f}, P \frac{d}{dt} \mathbf{g} \rangle = - \langle \frac{d}{dt} (P^T \mathbf{f}), \mathbf{g} \rangle$, for any matrix function $P(t) \in \mathbb{R}^{d \times d}$ which we use to symmetrise the second term in Eq. 2.87b. We get

$$S[\bar{\mathbf{x}} + \delta \mathbf{x}] = S[\bar{\mathbf{x}}(t)] + \langle \mathbf{j}, \delta \mathbf{x} \rangle + \frac{1}{2} \langle \delta \mathbf{x}, \mathcal{H} \delta \mathbf{x} \rangle + O(\delta \mathbf{x}^3) \quad (2.88)$$

where

$$\mathbf{j}^{[\bar{\mathbf{x}}]}(t) = \frac{\partial L}{\partial \mathbf{x}} - \frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{x}}} \quad (2.89)$$

$$\mathcal{H}^{[\bar{\mathbf{x}}]} = -\frac{1}{D} \frac{d}{dt^2} + 2A \frac{d}{dt} + B \quad (2.90)$$

and $A_{ij} = \frac{\partial^2 L}{\partial x_i \partial \dot{x}_j}$, $B_{ij} = \frac{\partial^2 L}{\partial x_i \partial x_j} - \frac{d}{dt} \frac{\partial^2 L}{\partial \dot{x}_j \partial \dot{x}_i}$, where closed brackets indicate an anti-symmetrisation over indices. The matrix-functions A and B are, given the reference path $\bar{\mathbf{x}}$, functions of time. \mathcal{H} can be seen as the infinite-dimensional Hessian operator of the expansion. In this context however, we will refer to it as the precision operator. For brevity we will henceforth omit the superscripts for Eq. 2.89 and Eq. 2.90.

By completing the square, we can rewrite Eq. 2.88 as

$$S[\bar{\mathbf{x}} + \delta \mathbf{x}] = S[\bar{\mathbf{x}}] - \langle \mathcal{H}^{-1} \mathbf{j}, \mathcal{H}^{-1} \mathbf{j} \rangle + \frac{1}{2} \langle \delta \mathbf{x} + \mathcal{H}^{-1} \mathbf{j}, \mathcal{H}(\delta \mathbf{x} + \mathcal{H}^{-1} \mathbf{j}) \rangle \quad (2.91)$$

where the first two terms are constants, for a given reference path $\bar{\mathbf{x}}$. From Eq. 2.84 we thus see that in this expansion the variations $\delta \mathbf{x}$ is a Gaussian process, which is distributed as $\mathcal{N}(-\mathcal{H}^{-1} \mathbf{j}, \mathcal{H}^{-1})$. We write the measure of the process as $\mathbb{P}_G^{[\alpha]}$, with density

$P_G^{[\bar{\mathbf{x}}]}[\delta\mathbf{x}] \propto \exp(-\frac{1}{2}\langle\delta\mathbf{x} + \mathcal{H}^{-1}\mathbf{j}, \mathcal{H}(\delta\mathbf{x} + \mathcal{H}^{-1}\mathbf{j})\rangle) \sim \mathbb{P}_G^{[\alpha]}$. For systems with gradient dynamics $\mathbf{b} = -\nabla U$, the asymmetric term in Eq. 2.90 vanishes, and the form of the precision operator simplifies to

$$\mathcal{H} = -\frac{1}{D}\frac{d^2}{dt^2} + B(t). \quad (2.92)$$

If the reference path is a local instanton of the transition path ensemble, then it is a local minimiser of the Onsager-Machlup action. It therefore solves the Euler-Lagrange equation of Eq. 2.85, and $\mathbf{j} = 0$. In this case the Gaussian process simplifies to $P_G^{[\bar{\mathbf{x}}]}[\delta\mathbf{x}] \propto \exp(-\frac{1}{2}\langle\delta\mathbf{x}, \mathcal{H}\delta\mathbf{x}\rangle)$.

We now reintroduce the superscript for the precision operator. For a purely diffusive process $\mathbf{b} = 0$ the precision operator is

$$\mathcal{H}_0 = -\frac{1}{D}\frac{d}{dt^2}. \quad (2.93)$$

Using integration-by-parts it can be shown that $S_0[\delta\mathbf{x}] = \frac{1}{2}\langle\delta\mathbf{x}, \mathcal{H}_0\delta\mathbf{x}\rangle$, where S_0 is the stochastic action for the Brownian bridge process. Therefore Eq. 2.93 is the precision operator of the Brownian bridge process (or equivalently, of the unpinned diffusions, the Wiener process). Now, as $\mathcal{H}^{[\bar{\mathbf{x}}]} = \mathcal{H}_0 + 2A\frac{d}{dt} + B$ we can write

$$\frac{d\mathbb{P}_G^{[\bar{\mathbf{x}}]}}{d\mathbb{P}_{\mathbf{B}}}[\delta\mathbf{x}] = (\mathcal{Z}^{[\bar{\mathbf{x}}]}/\mathcal{Z}^{\mathbf{B}})^{-1} \exp\left(-\frac{1}{2}\langle\delta\mathbf{x}, \mathcal{K}^{[\bar{\mathbf{x}}]}\delta\mathbf{x}\rangle\right) \quad (2.94)$$

where $\bar{\mathbf{x}}$ is a local instanton of the OM action, and where we have defined

$$\mathcal{K}^{[\bar{\mathbf{x}}]} = \mathcal{H}^{[\bar{\mathbf{x}}]} - \mathcal{H}_0 = 2A\frac{d}{dt} + B \quad (2.95)$$

which can be seen as a relative precision operator with respect to the precision operator \mathcal{H}_0 of the Brownian bridge process. In Eq. 2.94 $\mathcal{Z}^{[\bar{\mathbf{x}}]}$ and $\mathcal{Z}^{\mathbf{B}}$ are the ‘normalisation constants’ of $P_G^{[\bar{\mathbf{x}}]}$ and $P_{\mathbf{B}}$ respectively. These are divergent quantities by themselves, but their ratio is well-defined [33]. We call $\mathcal{Z}^{[\bar{\mathbf{x}}]}/\mathcal{Z}^{\mathbf{B}}$ a *regularised normalisation constant*, and in Appendix B we derive a method by which to compute them. However this result will not be necessary for the purposes of this chapter.

2.4.2 The Teleporter MCMC method

We will now use the results of the previous subsection to apply the TMC for sampling the transition path ensemble of Itô diffusion equations. Let $\mathbf{x}^{[\alpha]}, \alpha = 1, \dots, K$ be the local instantons of the system, given end-points conditions $\mathbf{X}(0) = \mathbf{x}_0$ and $\mathbf{X}(T) = \mathbf{x}_T$. Let $\mathcal{H}^{[\alpha]}$ be the precision operator of the variational expansion around $\mathbf{x}^{[\alpha]}$, and we define the

Gaussian probability measure $\mathbb{P}^{[\alpha]}$ over the TPE with density

$$P_G^{[\alpha]}[\mathbf{X}] = \mathcal{Z}^{[\alpha]} \exp\left(-\frac{1}{2}\langle \mathbf{X} - \mathbf{x}^{[\alpha]}, \mathcal{H}^{[\alpha]}(\mathbf{X} - \mathbf{x}^{[\alpha]}) \rangle\right) \sim \mathbb{P}_G^{[\alpha]}. \quad (2.96)$$

We construct the teleportation measure as

$$\mathbb{P}_T = \sum_{\alpha=1}^K w_\alpha \mathbb{P}_G^{[\alpha]} \quad (2.97)$$

where w_α are tuneable weights that must satisfy $\sum_{\alpha=1}^K w_\alpha = 1$. We must now compute the teleporter acceptance kernel Eq. 2.80. Firstly, we note that as $P_G^{[\alpha]}[\mathbf{X}]$ is an un-centred Gaussian with mean $\mathbf{x}^{[\alpha]}$, and can be factorised in terms of a centred density as

$$\begin{aligned} P_G^{[\alpha]}[\mathbf{X}] &= \mathcal{Z}^{[\alpha]} \times \exp\left(\frac{1}{2}\langle \mathbf{X}, \mathcal{H}^{[\alpha]}\mathbf{x}^{[\alpha]} \rangle + \frac{1}{2}\langle \mathbf{x}^{[\alpha]}, \mathcal{H}^{[\alpha]}\mathbf{X} \rangle - \frac{1}{2}\langle \mathbf{x}^{[\alpha]}, \mathcal{H}^{[\alpha]}\mathbf{x}^{[\alpha]} \rangle\right) \\ &\quad \times \exp\left(-\frac{1}{2}\langle \mathbf{X}, \mathcal{H}^{[\alpha]}\mathbf{X} \rangle\right) \end{aligned} \quad (2.98)$$

where the second factor can thus be seen as a Radon-Nikodym derivative, and is an instance of the Cameron-Martin theorem [117], which describes how infinite-dimensional Gaussian measures transform under translation. The density of $\mathbb{P}^{[\alpha]}$ with respect to $\mathbb{P}_{\mathbf{B}}$ is thus, using Eq. 2.95

$$\frac{d\mathbb{P}_G^{[\alpha]}}{d\mathbb{P}_{\mathbf{B}}} = (\mathcal{Z}^{[\alpha]}/\mathcal{Z}^{\mathbf{B}})^{-1} \exp\left(-\Psi^{[\alpha]}[\mathbf{X}]\right) \quad (2.99)$$

such that $P_G^{[\alpha]}[\mathbf{X}] = \frac{d\mathbb{P}_G^{[\alpha]}}{d\mathbb{P}_{\mathbf{B}}}[\mathbf{X}] P_{\mathbf{B}}[\mathbf{X}]$, and where

$$\Psi^{[\alpha]}[\mathbf{X}] = -\frac{1}{2}\langle \mathbf{X}, \mathcal{H}^{[\alpha]}\mathbf{x}^{[\alpha]} \rangle - \frac{1}{2}\langle \mathbf{x}^{[\alpha]}, \mathcal{H}^{[\alpha]}\mathbf{X} \rangle + \frac{1}{2}\langle \mathbf{x}^{[\alpha]}, \mathcal{H}^{[\alpha]}\mathbf{x}^{[\alpha]} \rangle + \frac{1}{2}\langle \mathbf{X}, \mathcal{K}^{[\alpha]}\mathbf{X} \rangle. \quad (2.100)$$

Using integration-by-parts we find

$$\Psi^{[\alpha]}[\mathbf{X}] = \frac{1}{2}\langle \mathbf{x}^{[\alpha]}, \mathcal{H}^{[\alpha]}\mathbf{x}^{[\alpha]} \rangle + \int_0^T \left\{ \frac{1}{D} \dot{\mathbf{x}}^{[\alpha]T} d\mathbf{X} + \mathbf{X}^T A^{[\alpha]} d\mathbf{X} + \frac{1}{2} \mathbf{X}^T B^{[\alpha]} \mathbf{X} dt \right\} \quad (2.101)$$

where the first term is a constant dependent on $\mathbf{x}^{[\alpha]}$ and the first and second terms in the integrand are stochastic Itô integrals. Substituting Eq. 2.100 into Eq. 2.80 we find

$$\begin{aligned} A_T[\mathbf{X}^{(n)}, \mathbf{X}'] &= \min \left\{ 1, \frac{P_{\mathbf{X}}[\mathbf{Y}] P_T[\mathbf{X}]}{P_{\mathbf{X}}[\mathbf{X}] P_T[\mathbf{Y}]} \right\} \\ &= \min \left\{ 1, \frac{P_{\mathbf{X}}[\mathbf{Y}] P_{\mathbf{B}}[\mathbf{X}] \sum_{\alpha=1}^K w_\alpha (\mathcal{Z}^{[\alpha]}/\mathcal{Z}^{\mathbf{B}})^{-1} e^{-\Psi^{[\alpha]}[\mathbf{X}]} }{P_{\mathbf{X}}[\mathbf{X}] P_{\mathbf{B}}[\mathbf{Y}] \sum_{\alpha=1}^K w_\alpha (\mathcal{Z}^{[\alpha]}/\mathcal{Z}^{\mathbf{B}})^{-1} e^{-\Psi^{[\alpha]}[\mathbf{Y}]}} \right\}. \end{aligned} \quad (2.102)$$

Finally, using $P_{\mathbf{X}}[\mathbf{X}] \propto e^{-\Phi[\mathbf{X}]} P_{\mathbf{B}}[\mathbf{X}]$ we find

$$A_T[\mathbf{X}, \mathbf{Y}] = \min \left\{ 1, \exp(\Phi[\mathbf{X}] - \Phi[\mathbf{Y}]) \frac{\sum_{\alpha=1}^K w_\alpha (\mathcal{Z}^{[\alpha]}/\mathcal{Z}^{\mathbf{B}})^{-1} e^{-\Psi^{[\alpha]}[\mathbf{X}]}}{\sum_{\alpha=1}^K w_\alpha (\mathcal{Z}^{[\alpha]}/\mathcal{Z}^{\mathbf{B}})^{-1} e^{-\Psi^{[\alpha]}[\mathbf{Y}]}} \right\}. \quad (2.103)$$

Equations 2.103 and 2.97, together with Eq. 2.78, where we use the pCN transition kernel $T_{\text{pCN}}[\mathbf{X} \rightarrow \mathbf{Y}]$ for the first term, completely specified the teleporter MCMC algorithm defined on the TPE. However, we must now discretise the algorithm to make it usable in numerics for the pCN method. As in Sec. 2.3.2, we expand sample paths in the KKL basis of the Brownian bridge process as Eq. 2.48. The resulting TMC algorithm will thus be defined on the dN -dimensional space of mode coefficients A , where N is the mode truncation order. Furthermore, though the distributions $\mathbb{P}^{[\alpha]}$ are Gaussian and it is possible to construct independence samplers for them, in what follows we will instead be using the results of Sec. 2.3.4 to construct simplified finite-dimensional measures.

For each $\alpha = 1, \dots, K$ we approximate the variational Gaussian measure $\mathbb{P}_G^{[\alpha]}$ as a multivariate Gaussian by expanding its precision operator $\mathcal{H}^{[\alpha]}$ in the KKL basis Eq. 2.45 of the Brownian bridge process. The resulting dL -dimensional multi-variate Gaussian has precision matrix

$$H_{ik,jl}^{[\alpha]} = \langle \mathbf{e}_i \phi_k, \mathcal{H}^{[\alpha]} \mathbf{e}_j \phi_l \rangle \quad (2.104)$$

for $i, j = 1, \dots, L$, $k, l = 1, \dots, d$, where, \mathbf{e}_k is a constant vector with one non-zero component $(e_k)_l = \delta_{kl}$ and $\phi_k(t)$ is given in Eq. 2.46. Though $H_{ik,jl}^{[\alpha]}$ is written with four indices, it should be seen as a square matrix, with two multi-indices i, k and j, l . As we discussed in Sec. 2.3.4, we would expect $H_{ik,jl}^{[\alpha]}$ to increasingly resemble the corresponding discretised precision matrix of the Brownian bridge process for high mode numbers k and l . That is, for some k_1 we have that

$$H_{ik,jl}^{[\alpha]} \approx \langle \mathbf{e}_i \phi_k, \mathcal{H}_0 \mathbf{e}_j \phi_l \rangle = \frac{1}{2D\lambda_k} \delta_{ij} \delta_{kl} \quad (2.105)$$

where $k, l > k_1$ and λ_k is given in Eq. 2.46. Using this fact, we construct an approximation of Eq. 2.104 up to dN dimensions as

$$\tilde{H}_{ik,jl}^{[\alpha]} = 2D\sqrt{\lambda_k}\sqrt{\lambda_l} \langle \mathbf{e}_i \phi_k, \mathcal{H}^{[\alpha]} \mathbf{e}_j \phi_l \rangle, \quad k, l \leq L \quad (2.106a)$$

$$\tilde{H}_{ik,jl}^{[\alpha]} = \delta_{ij} \delta_{kl}, \quad k, l > L \quad (2.106b)$$

where we have multiplied each entry of the matrix with a factor of $2D\sqrt{\lambda_k}\sqrt{\lambda_l}$ to normalise the diagonal variances. The approximation is accurate if $L \geq k_1$. This is equivalent to constructing a ‘grafted’ Gaussian process, where modes $k, l < L$ are distributed like the Gaussian process $\mathbb{P}_G^{[\alpha]}$, and modes $k, l > L$ are distributed like $\mathbb{P}_{\mathbf{B}}$. This defines a

dN -dimensional measure $\mathbb{P}_G^{[\alpha],N,L}$ on mode-space with density

$$\begin{aligned} P_G^{[\alpha],N,L}[A] &= \frac{1}{\sqrt{(2\pi)^{dN} \det(\tilde{H}^{[\alpha]})}} \\ &\times \exp \left(-\frac{1}{2} \sum_{i,j=1}^d \sum_{k,l=1}^N (A_{ik} - A_{ik}^{[\alpha]}) \tilde{H}_{ik,jl}^{[\alpha]} (A_{jl} - A_{jl}^{[\alpha]}) \right) \end{aligned} \quad (2.107)$$

where $A_{ik}^{[\alpha]} = \langle \phi_k, x_i^{[\alpha]} \rangle$ are the mode coefficients of the instantons $\mathbf{x}^{[\alpha]}$. Due to the fact that $\mathcal{H}^{[\alpha]}$ is unit-diagonal for $k, l > L$, the determinant can be computed using only the first $dL \times dL$ sub-matrix. Constructively, we can sample the stochastic paths of this process using the expansion

$$\mathbf{G}^{[\alpha]}(t) = \mathbf{x}^{[\alpha]}(t) + \sqrt{2D} \sum_{k=1}^N \mathbf{Z}_k^{[\alpha]} \sqrt{\lambda_k} \phi_k(t) \quad (2.108)$$

where $(\mathbf{Z}_1^{[\alpha]}, \dots, \mathbf{Z}_N^{[\alpha]})$ are sampled from the dN -dimensional multi-variate Gaussian $\mathcal{N}(0, \tilde{H}^{[\alpha]})$. Note however due to $\tilde{H}^{[\alpha]}$ being diagonal for $k > L$, we have that $Z_{ik}^{[\alpha]}, k > L$ are independent and are all distributed as $\mathcal{N}(0, 1)$. The result of all of the above is that we have now have a finite-dimensional transporter measure

$$\mathbb{P}_T^{N,L} = \sum_{\alpha=1}^K w_\alpha \mathbb{P}_G^{[\alpha],N,L}. \quad (2.109)$$

The corresponding teleporter acceptance kernel is now

$$A_T[A, A'] = \min \left\{ 1, e^{\Phi[\mathbf{X}(t;A)] - \Phi[\mathbf{X}(t;A')]} \frac{\sum_{\alpha=1}^K w_\alpha (\det(\tilde{H}^{[\alpha]}))^{-1} e^{-\tilde{\Psi}^{[\alpha]}[A]}}{\sum_{\alpha=1}^K w_\alpha (\det(\tilde{H}^{[\alpha]}))^{-1} e^{-\tilde{\Psi}^{[\alpha]}[A']}} \right\}. \quad (2.110)$$

where

$$\tilde{\Psi}^{[\alpha]}[A] = C^{[\alpha]} + \sum_{i=1}^d \sum_{k=1}^N A_{ik}^{[\alpha]} A_{ik} + \sum_{i,j=1}^d \sum_{k,l=1}^L A_{ik} \tilde{K}_{ik,jl}^{[\alpha]} A_{jl} \quad (2.111)$$

where we have defined

$$C^{[\alpha]} = \frac{1}{2} \sum_{i,j=1}^d \sum_{k,l=1}^N A_{ik}^{[\alpha]} \tilde{H}_{ik,jl}^{[\alpha]} A_{jl}^{[\alpha]} \quad (2.112a)$$

$$\tilde{K}_{ik,jl}^{[\alpha]} = \tilde{H}_{ik,jl}^{[\alpha]} - \delta_{ij} \delta_{kl}. \quad (2.112b)$$

where $\tilde{K}_{ik,jl}^{[\alpha]}$ are discretised relative precision matrices with respect to the precision matrix of the Brownian bridge process. Note that $A^{[\alpha]}$ and $C^{[\alpha]}$ are constant, and only need to be evaluated once given the instantons $\mathbf{x}^{[\alpha]}$. Due to the fact that $\mathbb{P}_G^{[\alpha],N,L}$ are spectrally grafted

in the manner described above, the complexity of sampling teleporter paths using Eq. 2.107 and Eq. 2.108 scales as $O(L^2)$ and not $O(N^2)$. Furthermore in practical applications, L is often a fixed value much smaller than N . As the value of L does not affect the target measure we sample, we can in general choose $L < k_{\parallel}$. Numerically, this also means that Eq. 2.111 will not diverge as $N \rightarrow \infty$, and nor will $\det(\tilde{H}^{[\alpha]})$.

This completes the mathematical inventory necessary for the transporter MCMC defined on the transition path ensemble. In the next subsection we will summarise the algorithm and present benchmark results.

2.4.3 Numerical algorithm

We can summarise the teleport MCMC method for sampling the transition path ensemble as follows:

1. Choose the mode truncation parameter N .
2. Find the K local instantons of the Onsager-Machlup action Eq. 2.9

$$\mathbf{x}^{[\alpha]} = \operatorname{argmin}_{\mathbf{x}} S_{\text{OM}}[\mathbf{x}] \quad (2.113)$$

where $\alpha = 1, \dots, K$, and project them onto the KKL basis of the Brownian bridge process

$$A_{ik}^{[\alpha]} = \langle \phi_k, x_i^{[\alpha]} \rangle. \quad (2.114)$$

3. Choose the truncation parameter L for the teleporter measure, the teleportation probability p_T and the teleportation weights w_α , $\alpha = 1, \dots, K$.
4. Compute the discretised precision matrix $\tilde{H}^{[\alpha]}$ using Eq. 2.106 and relative precision matrix $\tilde{K}^{[\alpha]}$ using Eq. 2.112b for each $\alpha = 1, \dots, K$, and let

$$\tilde{w}_\alpha = (\det(\tilde{H}^{[\alpha]}))^{-1} w_\alpha \quad (2.115)$$

5. Choose an initial state $A^{(0)} \in \mathbb{R}^{d \times N}$.

6. Draw a random number $V \sim \text{Unif}([0, 1])$

- If $V > p_T$:
- (a) Generate a proposal state

$$A'_{ik} = \sqrt{1 - \kappa^2} A_{ik}^{(n)} + \kappa_{ik} Z_{ik} \quad (2.116)$$

for each $i = 1, \dots, d$ and $k = 1, \dots, N$, where Z_{ik} is sampled from $\mathcal{N}(0, 1)$ and κ_{ik} is the step-size matrix.

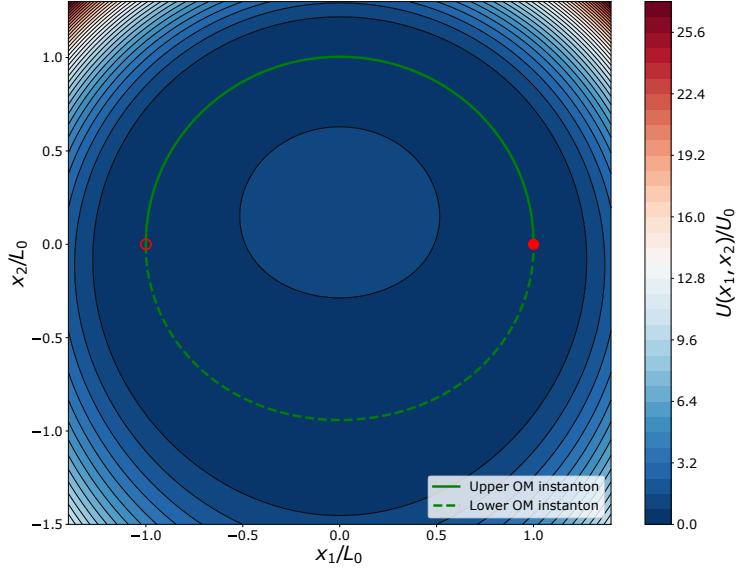


Figure 2.8: The switch potential $U(x_1, x_2)$, defined in Appendix A.2. There are two main transition channels between the fixed end-points $\mathbf{x}_0/L_0 = (-1, 0)$ (hollow red circle) and $\mathbf{x}_T/L_0 = (1, 0)$ (filled red circle). The potential has a higher curvature in the upper channel relative to the lower one, which can be seen by the wider contour regions in the lower half of the plot. The upper (green solid line) and lower (green dashed line) local Onsager-Machlup instantons were computed at temperature $\theta/\theta_0 = 0.1$ and duration $T/T_0 = 3$.

- (b) Draw a random number U from $\text{Unif}([0, 1])$.
 - If $U < A_{\text{pCN}}[A^{(n)}, A']$, then set $A^{(n+1)} = A'$.
 - Otherwise set $A^{(n+1)} = A^{(n)}$.
- Otherwise:
 - (a) Draw a random integer γ from $\{1, \dots, K\}$, with probabilistic weights $\{\tilde{w}_1, \dots, \tilde{w}_K\}$.
 - (b) Generate a proposal teleportation state

$$A' = Z^{[\gamma]} \quad (2.117)$$

where $Z^{[\gamma]}$ is drawn from the multi-variate Gaussian $\mathcal{N}(0, \tilde{H}^{[\alpha]})$.

- (c) Draw a random number U from $\text{Unif}([0, 1])$.
 - If $U < A_{\text{T}}[A^{(n)}, A']$, then set $A^{(n+1)} = A'$.
 - Otherwise set $A^{(n+1)} = A^{(n)}$.

7. Repeat step 6.

As a benchmark problem to test the TMC algorithm on, we have defined a system we call the *switch*. We consider the motion of a particle in a two-dimensional energy

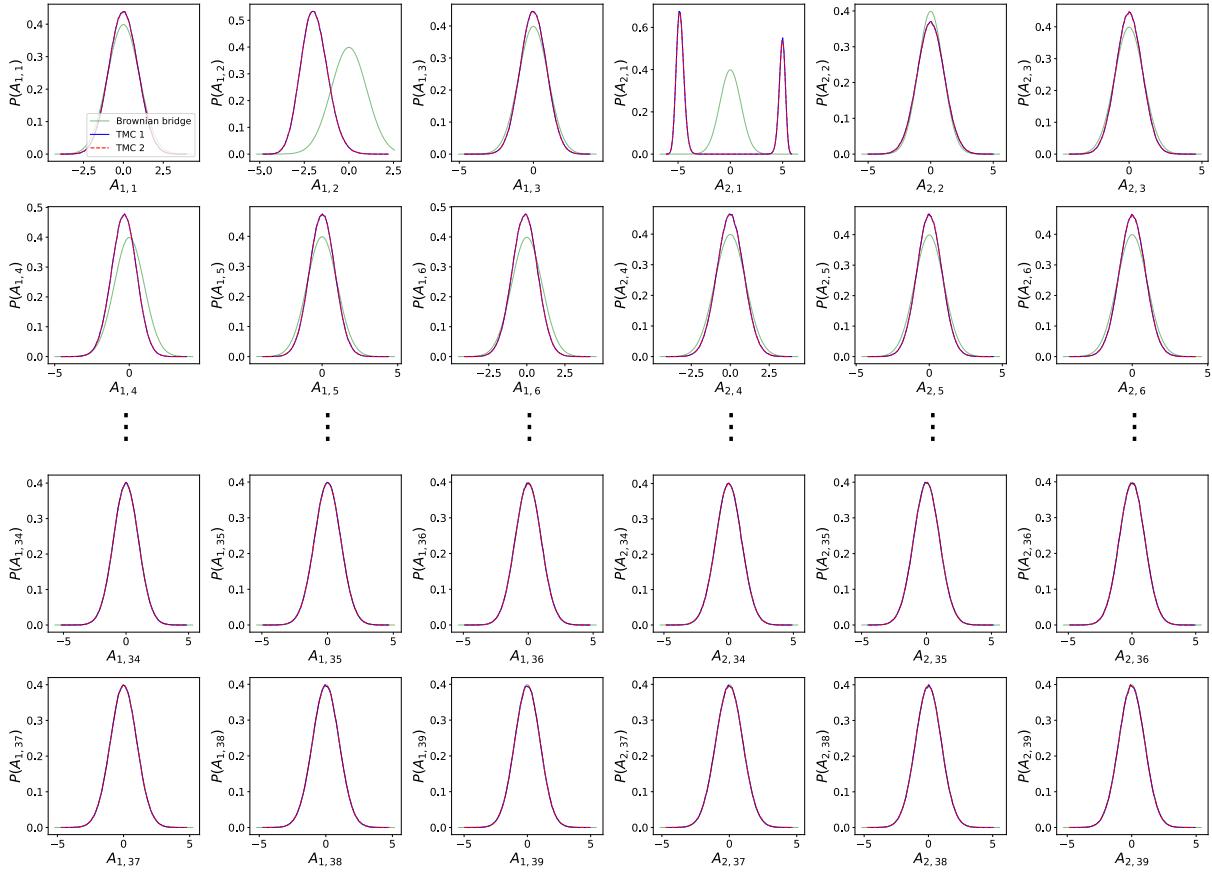


Figure 2.9: The marginalised mode distributions of two simulations of the TPE of the switch system using the TMC method with $\theta/\theta_0 = 3.36$ and $T/T_0 = 3$ and $M = 10^8$ samples, and the modes of the Brownian bridge process (green solid lines). The first (TMC 1, blue solid lines) was computed using equal teleporter weights $w_+ = w_- = 0.5$ for both channels, and the second (TMC 2, dashed red lines) was biased in favour of the upper channel $w_+ = 0.9, w_- = 0.1$. We see that the mode distributions of the two simulations agree. The mode distribution of $A_{2,1}$ reflects the fact that transition paths concentrate along the upper and lower transition channels.

landscape given by a potential function. The potential $U(\mathbf{x})$ is a deformed Mexican hat, with a maximum at the origin and a manifold of minima on the circle of radius L_0 around the origin. See Fig. 2.8 for a plot of $U(\mathbf{x})$, and Appendix A.2 for a detailed specification of the model. We consider the TPE of paths of duration T , which start at $\mathbf{x}_0/L_0 = (-1, 0)$ and end at $\mathbf{x}_T/L_0 = (1, 0)$. This ensemble features two competing transition channels, namely along the upper and lower semi-circle, which we denote by Γ^+ and Γ^- ; by design the potential along Γ^+ is narrower as compared to along Γ^- .

We used the TMC to sample the TPE of the switch system. For temperature $\theta/\theta_0 = 3.36$ and duration $T/T_0 = 3$, we computed the instantons $\mathbf{x}^{[+]}$ and $\mathbf{x}^{[-]}$ lying in the upper and lower transition channels respectively. We then followed the steps outlined above to implement the TMC. Let w_+ and w_- be the teleportation weights of the teleporters of $\mathbf{x}^{[+]}$ and $\mathbf{x}^{[-]}$ respectively. We ran two simulations, with different choice of teleportation

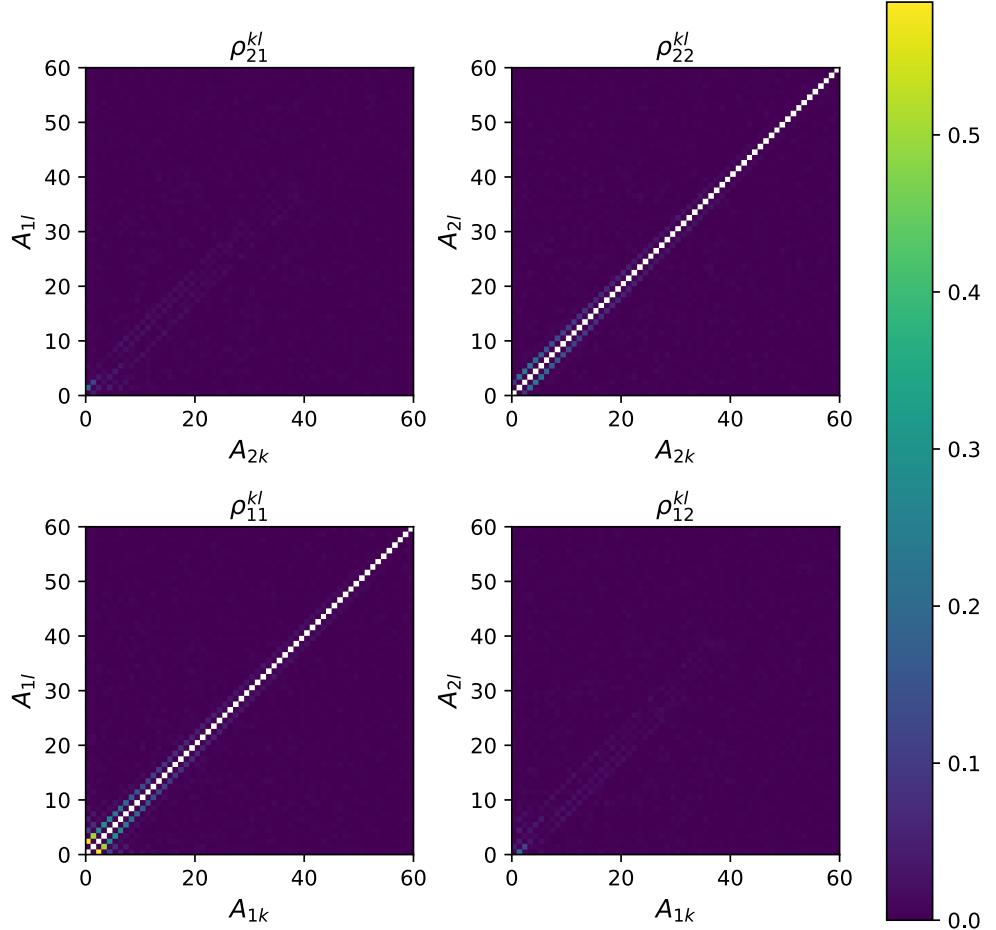


Figure 2.10: The absolute normalised covariances $\rho_{ik,jl} = \langle A_{ik} A_{jl} \rangle / \sqrt{\langle A_{ik}^2 \rangle \langle A_{jl}^2 \rangle}$ of the first 60 modes of the transition paths of the switch system. As $\rho_{ik,ik} = 1$, $i = 1, 2$, $k = 1, \dots, N$ by definition, we have excluded the diagonals from the plot. We sampled the TPE using the TMC algorithm, with parameters $\theta/\theta_0 = 3.36$, $T/T_0 = 3$ and $N = 200(T/T_0)$ with $M = 10^8$ iterations. We see that mode rapidly de-correlate for high mode numbers.

weights. In scenario 1 we used equal weights $w_+ = w_- = 0.5$, and in scenario 2 we biased in favour of the upper channel $w_+ = 0.9$ and $w_- = 0.1$. The two simulation scenarios lead to the same mode distributions. The results are depicted in Fig. 2.9 and Fig. 2.10. We also see the same spectral band separation we discussed in Sec. 2.3.4. Further discussion and analysis of the switch system will be done in Ch. 3, where it will be used to probe the relation between transition channels and diffusivity. The results of Ch. 3 will also show that the TMC is robust at low diffusivity regimes, sampling both transition channels, where standard MCMC methods (including the pCN) are intractable.

2.5 Conclusion

In this chapter we applied some recent mathematical developments in the field of functional Markov chain Monte Carlo methods to study the transition paths of stochastic processes.

We began with an overview of probability theory over infinite-dimensional sample spaces, and showed that the formal notion of a path-space probability density can be understood rigorously using the classical Wiener measure. Using these mathematical foundations we then described the preconditioned Nicolson-Crank algorithm and used it to sample the transition path ensemble of a general Itô diffusion equation with additive noise.

A suitable reference Gaussian measure of a pinned Itô diffusion is the Brownian bridge process, and we thus expanded transition paths in the Kosambi-Karhunen-Loëve (KKL) basis of the Brownian bridge. We provided numerical and analytical evidence of a band separation in the KKL mode-space of a general Itô diffusion with additive noise. The modes of the KKL expansion decouple in the high-frequency band, and statistically behave indistinguishably from that of Brownian motion. The only non-trivial statistics of an Itô diffusion is thus encapsulated by the joint distributions of the lower-band modes. We used this band separation to construct an adapted step-size for the pCN algorithm, and we demonstrated improved autocorrelation times.

The main result of this chapter was the teleporter MCMC (TMC). In the most general sense, the TMC is a Metropolis-Hastings scheme that interpolates between a given MCMC scheme and an independence sampler, where we called the latter the teleportation measure, and is constructed so that the proposal states concentrate around the maxima of the target measure. Our main innovation was to define such a sampler on the transition path ensembles of Itô processes. Using a semi-classical expansion of the path-probability density, we constructed approximate Gaussian measures around the maxima of the path measure. These formed the basis of the teleportation measure, and the resulting algorithm allowed for sampling TPEs with multiple dominant transition channels.

There are a number of promising future directions. We showed some results that the band separation of Itô processes in KKL mode-space could be exploited to moderately improve autocorrelation times. However, further work can be undertaken to combine the KKL expansion with a multi-level MCMC method [129, 130, 131, 132]. Briefly, we can coarse-grain the system in a series of mode truncations $N_1 < N_2 < \dots N_{M_{\text{multi}}}$. By running the pCN at the lowest-dimensional chain, we can then upsample via the multi-level scheme to generate samples on the highest-dimensional chain. We would use the band structure of the KKL expansion to engineer upsampling proposals adapted to the system.

Finally, we can consider extending the pCN and TMC to study higher-dimensional settings. The purview of this chapter was the sampling of stochastic transition paths $\mathbf{X}(t)$. Here, $t \in [0, T]$ is an index over a time domain, and the trajectory itself is a map $t \mapsto \mathbf{X}(t)$. In general, we could consider stochastic processes with two or more indices $(t, \mathbf{u}) \mapsto \mathbf{X}(t, \mathbf{u})$, where $\mathbf{u} \in M$ is here a spatial coordinate, that takes values in some n -dimensional topological space M . Examples of such systems are the stochastic field theories prominent in soft matter physics [133, 134, 135, 136, 137]. We will provide a

brief sketch of the necessary steps required to sample the transition path ensemble of such systems. As a first example we could consider *Model A* [138]

$$\partial_t \phi = -\mu \frac{\delta F}{\delta \phi} + \sqrt{2D}\xi \quad (2.118)$$

where $\phi = \phi(t, \mathbf{r})$ is a scalar field defined on M , $\frac{\delta}{\delta \phi}$ is the functional derivative and ξ is space-time white noise, and F is a free energy functional. A typical choice of F is the ϕ^4 free energy

$$F[\phi] = \int_M \left(\frac{a}{2} \phi^2 + \frac{b}{4} \phi^4 + \frac{\kappa}{2} (\nabla \phi)^2 \right) d\mathbf{r}. \quad (2.119)$$

Removing the ϕ^4 term leads to the *Ohrnstein-Uhlenbeck* (OU) [139] field equation

$$\partial_t \phi^{\text{OU}} = -\mu(a\phi^{\text{OU}} - \nabla^2 \phi^{\text{OU}}) + \sqrt{2D}\xi. \quad (2.120)$$

Now, if we expand paths in a Fourier basis as $\phi \propto \sum_{\mathbf{q}} \phi_{\mathbf{q}}^{\text{OU}} e^{i\mathbf{q} \cdot \mathbf{r}}$, we get decoupled stochastic equations of motion for each Fourier mode

$$\partial_t \phi_{\mathbf{q}}^{\text{OU}} = -\mu(a + \kappa q^2) \phi_{\mathbf{q}}^{\text{OU}} + \sqrt{2D}\xi. \quad (2.121)$$

We see that the Fourier modes of the ϕ^{OU} process behave like decoupled 1-dimensional OU processes. Now, it is clear that the OU field is a form of Gaussian process. If we presume that the field-probability measure of a general Model A system has a density with respect the OU field, it would be possible to construct a pCN scheme analogous to the ones described in this chapter by expanding ϕ in the KKL basis of ϕ^{OU} . The latter would require a generalised KKL expansion for fields, which have been discussed in [140], as well as a closed-form expression for the KKL basis of OU bridge processes, which were given here [141]. Finally, we note that such a generalisation of our MCMC methods would be applicable to the general class of continuum systems we treat in the second part of this thesis, if we include stochasticity in the latter.

Chapter 3

Diffusivity dependence of transition paths

3.1 Introduction

The fluctuating dynamics of many physical, chemical and biological systems are commonly modelled by stochastic differential equations expressed in Langevin or Itô forms [19, 20, 21, 22]. In such systems it is often of great interest to identify the typical pathways that stochastic paths take to transition from an initial to a final state, as for example in the nucleation of solids, the conformational changes in biomolecules, or shifts in ecological balance [9, 10, 13, 14, 15, 16, 17, 23, 24]. Typically, such transition paths cluster around multiple pathways in the space of configurations and the relative probability of one or the other of these pathways depends on the drift, the diffusivity, and the duration allowed for the transition to take place [25, 26, 27, 28]. As transitions are often rare events, direct simulations are not always practical and other means, analytical or numerical, are required to study them. Methods that allow for a full exploration of the space of transition pathways in stochastic dynamical systems, then, are of substantial theoretical and practical importance.

The theory of large deviations [2, 29, 30, 31], the topic of Ch. 1, provides an analytical method for obtaining transition pathways - instantons - in regimes dominated by the drift and for very long durations of path. Experimental systems, however, are typically not in a regime where the diffusivity is asymptotically low and durations are asymptotically long [32]. While the relevance of including finite-temperature fluctuations around the instanton [33] is increasingly recognized [34, 35, 36], the physical implications of these fluctuations are far from being understood.

In this chapter, we show that the competition between drift and diffusion in transition pathways can be studied using the semi-classical expansions of the path measure of the stochastic dynamics introduced in Sec. 2.4. We use a mixture of Gaussian measures to

approximate the path measure around its local instantons. This allows us to demarcate and transcend the boundaries of the low diffusivity regime, without recourse to sampling the TPE directly using MCMC methods. We demonstrate this explicitly for a two-dimensional overdamped mechanical system, with both conservative and non-conservative forces. For this system we uncover a counterintuitive phenomenon where typical transition paths do not concentrate around the most probable path, even at low-to-intermediate diffusivities where the Gaussian approximation is still valid. To validate our results numerically, and to study the TPE at high-diffusivity regimes, we use the teleporter MCMC (TMC) method developed in Ch. 2, that allows for simultaneous exploration of multiple transition pathways. We find excellent agreement between the semi-classical expansion and numerical results for a large range of diffusivities and path durations. The results presented in this chapter was the result of collaborative work published as a pre-print in [37], of which the author of this text was the primary contributor.

3.2 The transition path ensemble

The general class of systems in consideration is the d -dimensional overdamped Langevin equation with additive noise, which we write here again in Itô form as

$$d\mathbf{X} = \mu\mathbf{F}dt + \sqrt{2D}d\mathbf{W}. \quad (3.1)$$

This represents the stochastic displacement $d\mathbf{X}$ in a time interval dt of a particle with coordinate \mathbf{X} subject to a force field \mathbf{F} and Brownian displacements $\sigma d\mathbf{W}$, where \mathbf{W} is the Wiener process. The particle mobility is μ , the diffusion constant is $D = \mu/\beta$, and the temperature is θ with $\beta^{-1} = k_B\theta$, and k_B the Boltzmann constant. We are interested in realisations $\mathbf{X}(t)$ of Eq. 3.1 that are of duration T and have fixed terminii $\mathbf{X}(0) = \mathbf{x}_0$ and $\mathbf{X}(T) = \mathbf{x}_T$. These trajectories form the set of continuous paths that we have called the transition path ensemble (TPE), and denote as $E_{\mathbf{x}_0}^{\mathbf{x}}([0, T])$. We will study the temperature dependence of the TPE using both the TMC algorithm presented in Ch. 2, as well as the semi-classical approximation of the path-probability measure, which are developed in Sec. 3.3.

3.2.1 Model system

The methods we will develop in subsequent sections will be applied on the *switch* system, which was previously used in Sec. 2.4.3, and is defined in detail in Appendix A.2. The model consists of a particle in $d = 2$ dimensions in a potential force field $\mathbf{F} = -\nabla U(\mathbf{x})$, depicted in Fig. 2.8 and Fig. 3.1a. The potential $U(\mathbf{x})$ has two prominent transition channels, the upper and lower semi-circular channels which we denote as Γ^+ and Γ^-

respectively. The potential has a continuous manifold of minima along Γ^+ and Γ^- , such that the purely energetic cost of traversing either channel is the same. However, by design the potential has a larger curvature along Γ^+ relative to Γ^- . Therefore, the lower channel is wider than the upper channel, as can be seen in Fig. 2.8. As an additional model, we add a non-conservative circular force \mathbf{F}^a to the switch system. This force works as a bias in favour of the upper channel, by decreasing the energetic cost of traversing it, and increases the energetic cost of the lower channel. The models have been constructed so as to study the interplay between the deterministic drift and diffusivity.

3.3 Semi-analytical approximations of stochastic path measures

Here we derive an approximation of the path measure \mathbb{P}_X of the Itô diffusion process with additive noise. In Sec. 3.3.1 we construct a global approximation of \mathbb{P}_X using a weighted sum of the Gaussian distributions derived in Sec. 2.4.1, where the latter approximated the path measure around given local instantons. We then use this result in Sec. 2.4.1 to derive estimates of transition channel probabilities, which serve as a measure for the concentration of the TPE around a given transition channel. Our method is a combination of the semi-classical approach to study path-integrals [89, 90, 91, 92, 93, 94] and a standard technique in statistical data analysis [142, 143, 144, 145, 146], wherein a target probability measure is estimated as Gaussian mixture distribution.

We will begin by first giving some intuition by drawing an analogy in the one-dimensional case. For a one-dimensional probability density $p(x) = \mathcal{Z}^{-1} \exp(-V(x))$, where \mathcal{Z} is a normalization constant and where the potential $V(x)$ has well-separated relative minima x_α , $\alpha = 1, \dots, K$, we can approximate $p(x)$ around x_α using a Gaussian approximation

$$p(x) \approx \frac{1}{\mathcal{Z}} e^{-V(x_\alpha) - V''(x_\alpha)(x-x_\alpha)^2/2} =: \frac{\mathcal{Z}_\alpha}{\mathcal{Z}} e^{-V(x_\alpha)} p^{[\alpha]}(x) \quad (3.2)$$

with a normalised Gaussian distribution $p^{[\alpha]}(x) := \mathcal{Z}_\alpha^{-1} e^{-V''(x_\alpha)(x-x_\alpha)^2/2}$ and where $\mathcal{Z}_\alpha = \sqrt{2\pi/V''(x_\alpha)}$. Equation 3.2 is a local approximation of $p(x)$ around x_α . If $p(x)$ is highly peaked around its maxima (for example if $V(x)$ describes a Boltzmann distribution $V(x) = U(x)/(k_B\theta)$ at a low temperature θ), a global approximation of $p(x)$ is the Gaussian mixture

$$p(x) \approx \sum_{\alpha=1}^K \omega_\alpha p^{[\alpha]}(x) \quad (3.3)$$

where $\omega_\alpha = e^{-V(x_\alpha)} \mathcal{Z}_\alpha / \sum_{\gamma=1}^K e^{-V(x_\gamma)} \mathcal{Z}_\gamma$ are constants that weight the local Gaussian distributions. The weight ω_α is the size of the support of $p^{[\alpha]}$ relative to the other local

Gaussian. From Eq. 3.3 we see that the support of $p^{[\alpha]}$ is determined by an interplay of the ‘peak’ of the distribution $e^{-V(x_\alpha)}$, and the ‘width’, which is measured by the normalisation constant \mathcal{Z}_γ .

Equation 3.3 can be used to approximately evaluate any expectation value. In particular, the probability of being in well α (i.e. around x_α) is given by

$$\mathbb{P}(x \in \text{well } \alpha) = \langle \chi_\alpha \rangle = \int_{-\infty}^{\infty} dx \chi_\alpha(x) p(x) \approx \sum_{\beta=1}^K \omega_\beta \int_{-\infty}^{\infty} dx \chi_\alpha(x) p^{[\beta]}(x) \quad (3.4)$$

$$\approx w_\alpha \int_{-\infty}^{\infty} dx p^{[\alpha]}(x) = \omega_\alpha = \frac{e^{-V(x_\alpha)} \mathcal{Z}_\alpha}{\sum_{\gamma=1}^L e^{-V(x_\gamma)} \mathcal{Z}_\gamma}, \quad (3.5)$$

where the indicator function $\chi_\alpha(x)$ is 1 if x is in well α and zero otherwise, and where we assume that the potential wells of $V(x)$ are well-separated so that $\chi_\alpha(x)p^{[\beta]}(x)$ is negligibly small whenever $\alpha \neq \beta$. In the following we apply the same steps as above to the case of the TPE.

3.3.1 The Gaussian mixture approximation

We use the quadratic expansion of the stochastic action derived in Sec. 2.4.1 to construct an approximate probability measure over the TPE. Let $\mathbf{x}^{[\alpha]}, \alpha = 1, \dots, K$ be the local instantons of a given Langevin system. For each local instanton we have the Gaussian measures $\mathbb{P}_G^{[\alpha]}$ with mean $\mathbf{x}^{[\alpha]}$ and precision operators $\mathcal{H}^{[\alpha]}$, which locally approximates the path measure around $\mathbf{x}^{[\alpha]}$. We use these local approximants to construct a Gaussian mixture measure with density

$$\tilde{P}_{\mathbf{X}}[\mathbf{X}] = \sum_{\alpha=1}^K \omega_\alpha \exp \left(-\frac{1}{2} \langle \mathbf{X} - \mathbf{x}^{[\alpha]}, \mathcal{H}^{[\alpha]}(\mathbf{X} - \mathbf{x}^{[\alpha]}) \rangle \right) \quad (3.6)$$

with corresponding measure $\tilde{P} \sim \tilde{P}$. We now impose that $\tilde{P}_{\mathbf{X}}[\mathbf{x}^{[\alpha]}] = P_{\mathbf{X}}[\mathbf{x}^{[\alpha]}]$ for each $\alpha = 1, \dots, K$. We get

$$\omega_\alpha = \frac{e^{-S_{\text{OM}}[\mathbf{x}^{[\alpha]}]} (\mathcal{Z}^{[\alpha]} / \mathcal{Z}_B)}{\sum_{\gamma=1}^K e^{-S_{\text{OM}}[\mathbf{x}^{[\gamma]}]} (\mathcal{Z}^{[\gamma]} / \mathcal{Z}_B)} \quad (3.7)$$

where $\mathcal{Z}^{[\alpha]}$ and \mathcal{Z}_B formally represent the normalisation constants for the path-densities for the Itô process and Brownian bridge process respectively. Whilst divergent on their own, their ratio $\mathcal{Z}^{[\alpha]} / \mathcal{Z}_B$ is finite, and we derive a formula with which to compute them in Appendix B. The measure $\tilde{P}_{\mathbf{X}}$ serves as an approximation of the true path-measure $P_{\mathbf{X}}$. The underlying assumption is that $P_{\mathbf{X}}$ concentrates around its local instantons. In sufficiently low diffusivity-regimes, $\tilde{P}_{\mathbf{X}}$ will be a good approximation for $P_{\mathbf{X}}$.

The sequences of steps taken to construct Eq. 3.6 is standard technique from statistical

data analysis [142, 143, 144, 145, 146], where we have applied it here in the infinite-dimensional path-space setting.

3.3.2 Approximations of transition channel probabilities

We now derive estimates of transition channel probabilities using the Gaussian mixture approximation of the TPE $E_{\mathbf{x}_0}^{\mathbf{x}}([0, T])$. Let $E^{[\alpha]} \subset E_{\mathbf{x}_0}^{\mathbf{x}}([0, T])$, $\alpha = 1, \dots, J$, which are disjoint open sets in the TPE, be the J transition channels under consideration. We define

$$\rho^{[\alpha]}(\theta, T) = \mathbb{P}_{\mathbf{x}}[E^{[\alpha]}] \quad (3.8)$$

which is the probability of observing a transition path $\mathbf{X} \in E_{\alpha} \subset E_{\mathbf{x}_0}^{\mathbf{x}}([0, T])$. The quantity $\rho^{[\alpha]}(\theta, T)$ can be seen as an observable, and the expectation on the right-hand side of Eq. 3.8 will be computed by sampling the TPE using MCMC methods. However, here we will also approximate $\rho^{[\alpha]}(\theta, T)$ using the Gaussian mixture measure as

$$\rho^{[\alpha]}(\theta, T) \approx \rho_G^{[\alpha]} \equiv \tilde{\mathbb{P}}_{\mathbf{x}}[E^{[\alpha]}] \quad (3.9)$$

where we have defined the approximate transition channel probability $\rho_G^{[\alpha]}$.

At sufficiently low temperatures we can assume that $\mathbb{P}_{\mathbf{x}}(\cup_{\alpha} E^{[\alpha]}) \approx 1$, i.e. approximately all stochastic paths travel via one of the transition channels. We also assume that the transition channels concentrate around the local instantons of the path measure. That is, $J = K$, where K is the number of local instantons, and that $\mathbf{x}^{[\alpha]} \in E_{\alpha}$. Furthermore, we assume that each local Gaussian approximate measure $\mathbb{P}_G^{[\alpha]}$ lacks support on transition channels other than $E^{[\alpha]}$, that is $\mathbb{P}_G^{[\alpha]}(U_{\gamma \neq \alpha} E^{[\gamma]}) \approx 0$. Under these assumptions $\rho^{[\alpha]}(\theta, T)$ is well-approximated by $\rho_G^{[\alpha]}$, and we have

$$\rho_G^{[\alpha]}(\theta, T) = w_{\alpha} = \frac{e^{-S_{\text{OM}}[\mathbf{x}^{[\alpha]}]} (\mathcal{Z}^{[\alpha]} / \mathcal{Z}_{\mathbf{B}})}{\sum_{\gamma=1}^K e^{-S_{\text{OM}}[\mathbf{x}^{[\gamma]}]} (\mathcal{Z}^{[\gamma]} / \mathcal{Z}_{\mathbf{B}})} \quad (3.10)$$

According to Eq. 3.10 the relative channel probabilities are determined by an interplay between the instanton probabilities, as quantified by $e^{-S_{\text{OM}}[\mathbf{x}^{[\alpha]}]}$, and the amplitudes of the Gaussian fluctuations around the local instantons, measured by $\mathcal{Z}^{[\alpha]} / \mathcal{Z}_{\mathbf{B}}$. We see that it is possible for a transition channel to be dominant despite its local instanton having a low probability, if the ‘width’ $\mathcal{Z}^{[\alpha]} / \mathcal{Z}_{\mathbf{B}}$ compensates for it. It is instructive to compare $\rho_G^{[\alpha]}(\theta, T)$ with another estimator

$$\rho_I^{[\alpha]}(\theta, T) = e^{-S_{\text{OM}}[\mathbf{x}^{[\alpha]}]} / \sum_{\gamma} e^{-S_{\text{OM}}[\mathbf{x}^{[\gamma]}]} \quad (3.11)$$

in which only the instanton probabilities are retained.

To use Eq. 3.10 in practice, we retrieve the instantons $\mathbf{x}^{[\alpha]}$ using a Ritz variational method as discussed in Ch. 1. We subsequently evaluate the regularised normalisation constants $\mathcal{Z}^{[\alpha]}/\mathcal{Z}_{\mathbf{B}}$ using the method described in Appendix B. Note that the instantons $\mathbf{x}^{[\alpha]}$ and the regularised normalisations $\mathcal{Z}^{[\alpha]}/\mathcal{Z}_{\mathbf{B}}$ are both dependent on θ and T .

3.4 Results

We now consider the transition behavior of the 2D system depicted in Fig. 3.1a. For a range of temperatures θ and total transition times T we first generate ensembles of 10^8 sample transition paths per tuple (θ, T) using the TMC. Let $\tau_D(\theta) = L^2/(\mu k_B \theta)$, which is the diffusive time-scale at temperature θ . We also introduce fixed reference temperature and time-scales $\theta_0 = U_0/k_B$ and $T_0 = \tau_D(\theta_0)$, where U_0 is the energetic well-depth of the potential. Our parameter range is such that $T \ll \tau_D$, for each temperature θ in the range considered. Each sampled ensemble thus describes a rare transition event. Let $\rho^+(\theta, T)$ and $\rho^-(\theta, T)$ be the upper and lower transition channel probabilities respectively, satisfying $\rho^+(\theta, T) + \rho^-(\theta, T) = 1$. Furthermore, let $\mathbf{x}^+(t; \theta, T)$ and $\mathbf{x}^-(t; \theta, T)$ be the upper and lower local Onsager-Machlup instantons respectively at temperature θ and duration T , from which we also compute the approximate channel probability $\rho_G^+(\theta, T)$.

For a total transition time $T/T_0 = 3$, and for each of the two temperatures $\theta/\theta_0 = 0.047$ and $\theta/\theta_0 = 0.004$, we show 50 randomly chosen TMC sample paths in Fig. 3.1a. We observe that while for the higher temperature the paths are evenly distributed between the two channels, for the lower temperature the lower channel is preferred. In Fig. 3.1b we show the numerical evaluation of $\rho^+(\theta, T)$ using the TMC as a function of both θ and T . Consistent with the $\theta/\theta_0 = 0.047$ data from Fig. 3.1a, we observe that for large enough temperature $\rho^+(\theta, T) \approx 1/2$ (white region), so that upper and lower channel are equally probable. That at large temperature the asymmetry in U becomes irrelevant for the TPE is expected, as in this limit the random force in Eq. 3.1 dominates over the deterministic force. As θ is decreased, the channel around Γ^- becomes dominant, so that $\rho^+(\theta, T) \rightarrow 0$ (blue region in Fig. 3.1b, c.f. $\theta/\theta_0 = 0.004$ data in panel (a)). The exact temperature at which the crossover from the diffusivity-dominated regime to the drift-dominated regime occurs decreases with increasing T ; this is clearly seen in Fig. 3.1c where vertical sections of panel (b) are shown for several values of T .

We now compare our numerical TMC results for $\rho^+(\theta, T)$ with the Gaussian mixture approximation ρ_G^+ . Figure 3.1a shows that this approximation is valid in the low-temperature regime (plus signs). This is consistent with the assumptions underlying the Gaussian approximation, as we expect the probability distribution in path space to be dominated by the neighborhoods of the local instantons only for sufficiently low temperature. As Fig. 3.1c shows, $\rho_G^+(\theta, T)$ quantitatively captures the beginning of the crossover from drift-dominated

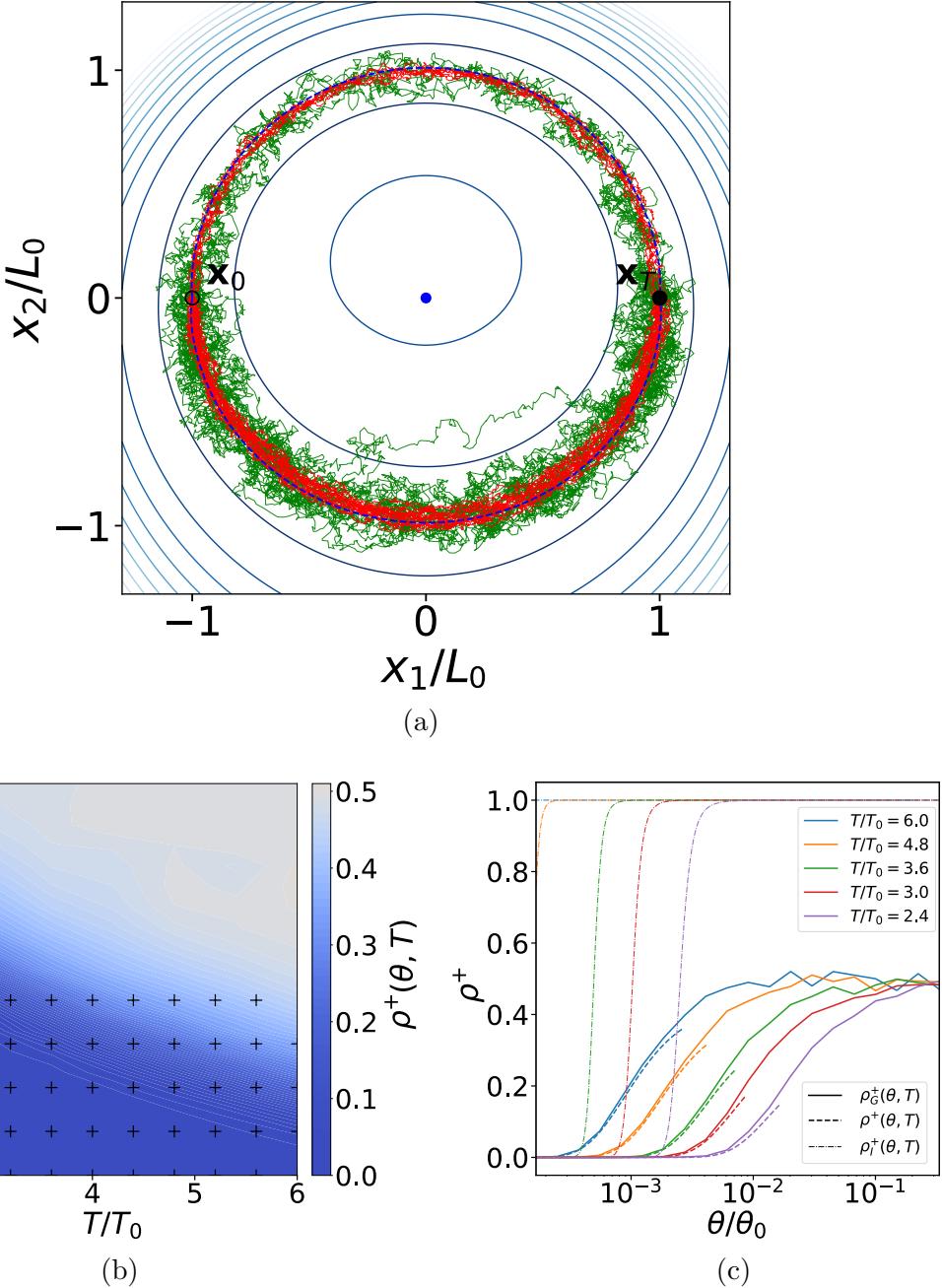


Figure 3.1: Diffusivity-dependence of the transition path ensemble for the conservative switch system. Panel (a) shows 50 stochastic trajectories sampled using the TMC method for overdamped dynamics in the potential $U(x)$ (see Appendix A.2). The dashed blue (green) lines are upper (lower) instantons between initial (circle, x_0) and final (filled circle, x_T) points. Upper and lower channels are equally populated at temperature $\theta/\theta_0 = 0.047$ (green) but the lower channel is preferred at the lower temperature $\theta/\theta_0 = 0.004$ (red). Trajectories of duration $T/T_0 = 3$ are sampled with $N = 200(T/T_0)$ modes. Panel (b) is a pseudocolor plot quantifying the variation of the upper channel probability ρ_G^+ with temperature and duration, as obtained from TMC. The plus signs show regions where the Gaussian mixture approximation ρ_G^+ , defined in Eq. 3.10, is within a 5% margin of error of the simulated value. The red and green dots correspond to the same color-coded simulations in panel (a). Panel (c) compares the upper channel probability $\rho^{[a]}(\theta, T)$, computed using the TMC, with the Gaussian mixture and instanton approximations to the upper channel probability as function of temperature for a range fixed durations of path T .

to diffusivity-dominated transition behaviour for all values of T considered.

For capturing this θ -dependent crossover, the prefactors $\mathcal{Z}^+ = \mathcal{Z}^{[1]}$, $\mathcal{Z}^- = \mathcal{Z}^{[2]}$ in Eq. 3.10 are essential. This becomes apparent by considering $\rho_I^+(\theta, T)$, which only depends on the relative probabilities of the two local instantons. In Fig. 3.1c we see that for high enough temperatures $\rho_I^+(\theta, T) \approx 1$, meaning $S_{\text{OM}}[\mathbf{x}^+(t; \theta, T)] < S_{\text{OM}}[\mathbf{x}^-(t; \theta, T)]$ [109]. This limit is understood by comparing the two terms in the action Eq. 2.85a. While the first term scales as $1/\theta$, the second term is independent of θ ; for fixed T and large enough θ the second term thus dominates the action. This second term is smaller for the channel around Γ^+ than for the channel around Γ^- , because the former channel is narrower leading to a smaller value of $\nabla \cdot \mathbf{F}$. As θ is decreased for fixed T the first term in Eq. 2.85a becomes dominant. Figure 3.1c shows that this leads to a crossover to $\rho_I^+(\theta, T) \approx 0$, meaning $\mathbf{x}^-(t; \theta, T)$ becomes more probable than $\mathbf{x}^+(t; \theta, T)$. While this low-temperature limit is consistent with the numerical results, the temperature at which we observe the crossover in $\rho_I^+(\theta, T)$ is smaller as compared to $\rho^+(\theta, T)$. For example, we see in Fig. 3.1c that for $T/T_0 = 2.4$ the crossover of $\rho_I^+(\theta, T)$ is at $\theta/\theta_0 < 10^{-2}$, whereas the crossover for $\rho^+(\theta, T)$ occurs at $\theta/\theta_0 > 10^{-2}$. In particular this implies that for $\theta/\theta_0 = 10^{-2}$ the most probable path goes along Γ^+ , while most transition paths go along Γ^- . This highlights that even at intermediate-to-low temperatures, where the Gaussian mixture approximation $\rho_G^+(\theta, T)$ is already valid, the probabilities of the local instantons alone are insufficient to obtain the actual transition behaviour. Instead it is the prefactors \mathcal{Z}^\pm in Eq. 3.10 that dominate the crossover behaviour in Fig. 3.1b; these Gaussian normalisation constants are, in a sense, an entropic contribution, as they measure the effective volume in path space of the support around the respective local instanton. Even though for $T/T_0 = 2.4$, $\theta/\theta_0 = 10^{-2}$ the instanton $\mathbf{x}^+(t; \theta, T)$ is more probable than $\mathbf{x}^-(t; \theta, T)$, this is more than offset by the larger number of paths that behave similar to $\mathbf{x}^-(t; \theta, T)$.

For the non-gradient forms of the drift, the prefactors \mathcal{Z}^\pm can also drive the crossover behaviour of the system, as we show now by adding a force of strength η that acts perpendicular to the radius vector in the clockwise direction. For positive force strength η , this non-conservative force biases towards the upper channel Γ^+ . In Fig. 3.2a we show numerical results for $\rho^+(\theta, T, \eta)$ as a function of η and θ for $T/T_0 = 3$. For small $\eta/f_{\text{eq}} \rightarrow 0$, where f_{eq} is the characteristic strength of the gradient force (see Appendix A.2 for more details), we recover the results from Fig. 3.1b and Fig. 3.1c. Thus at small but finite temperature the dominant transition channel is the one where particles travel against the weak non-conservative force. As η is increased to $\eta_c/f_{\text{eq}} \approx 0.00387$, we observe a crossover from Γ^- -channel dominated transitions to Γ^+ -channel dominated transitions in the low-temperature regime. This switch is also captured by the Gaussian approximation (plus signs in Fig. 3.2a). On the other hand, throughout the parameter regime considered in Fig. 3.2a, we find that $\rho_I^+(\theta, T, \eta) \approx 1$, meaning that the local instanton $\mathbf{x}^+(t; \theta, T)$ is

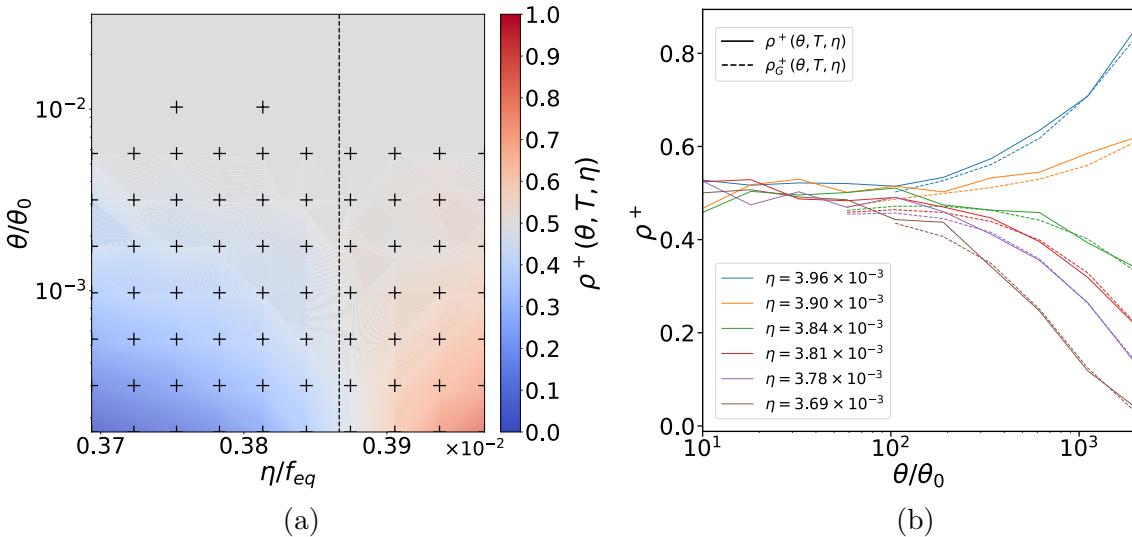


Figure 3.2: Diffusivity-dependence of the transition path ensemble for the non-conservative model system. Panel (a) is a pseudocolour plot of the probability of the upper channel $\rho^+(\theta, T, \eta)$, computed using the TMC, for the non-gradient system with $T/T_0 = 3$, as a function of the temperature θ and the circular force-strength η . The black plus signs show regions where the variational approximation $\rho_G^+(\theta, T, \eta)$, defined in Eq. 3.10, is within a 5% margin of error of the simulated value. The dashed line shows the crossover force strength $\eta_c/f_{eq} \approx 0.00387$, where f_{eq} is the characteristic strength of the gradient force. Panel (b) Compares the upper channel probability $\rho^{[\alpha]}(\theta, T)$ with the Gaussian mixture and instanton approximations to the upper channel probability as function of temperature for $T/T_0 = 3$ and a range of values for η_c .

always more probable than $\mathbf{x}^-(t; \theta, T)$ for finite η . This again highlights the relevance of considering Gaussian fluctuations around the instantons for determining the dominant transition pathway.

3.5 Conclusion

For a system with two competing transition pathways, we have studied how the dominant transition pathway depends on both the temperature and the total duration. To quantify the relative importance of the competing pathways, we have constructed semi-analytical approximators which are valid in the low-to-intermediate temperature regime. We have validated our approximators via comparison with a continuous-time MCMC method that is dimensionally robust and efficiently samples systems with multiple reactive pathways.

Our results show that even in the low-to-intermediate temperature regime the global instanton, or most probable path, itself is not sufficient to determine the dominant transition pathway. Rather, it is vital that fluctuations around this path be incorporated. This has a simple one-dimensional equivalent: For a probability density $p(x) \sim \exp(-V(x))$ for some potential $V(x)$ with relative minima x_α , the probabilistically most relevant minimum is not the global one, but that with the largest well probability, i.e. the x_α that

maximizes $P(\text{well around } x_\alpha) \sim e^{-V(x_\alpha)} \sqrt{2\pi/V''(x_\alpha)}$, where we use a quadratic Taylor approximation of V around x_α . The most probable well is thus determined by an interplay of $e^{-V(x_\alpha)}$ (which corresponds to the instanton probability $e^{-S_{\text{OM}}[\mathbf{x}^{[\alpha]}]}$) and $\sqrt{2\pi/V''(x_\alpha)}$ (which corresponds to the regularised normalisation constant $\mathcal{Z}^{[\alpha]}$).

In this chapter we considered a paradigmatic example system with two competing transition pathways. The method of instantons is an established technique in theoretical chemistry and statistical physics [43, 48, 54, 64, 110, 147, 148, 149, 150], and the method of Gaussian mixtures presented here scales as $O(d^2)$ with the number of degrees of freedom d . It is therefore feasible to apply the methods we developed here to more realistic many-particle systems to study e.g. nucleation pathways [43, 151, 152] or conformational rearrangements in macromolecules [105, 153, 154, 155].

Our quantification here of the finite-temperature breakdown of instanton theories is important for relating such theories to experiments, which are always at finite temperatures. Our insights into path-space probability distributions for diffusive stochastic dynamics, together with our MCMC method, will therefore be valuable for going beyond the regime of asymptotically low diffusivity in both large deviations theory [2, 64, 155] and the study of rare events [156].

Part II

Geometric mechanics of microstructured soft matter

Chapter 4

Introduction

Start with the ubiquity of the technique.

Then say that there is recently some study into the geometrisation of the cosserat rod (giusteri).

There is also extended literature on lie group mechanics (marsden), and integrators.

Here we build a theory that is a synthesis of all of these. We start with cosserat, then generalise, and then we also build lie integrators.

Continuum mechanics are well-studied in engineering, as well as cosserat systems with internal degrees of freedom. Simultaneously, there is a wide literature on geometric formulation of finite-dimensional systems, as well as on geometric integrators. The primary work of this text is to synthesise these two in a programmatic framework.

A wide range of structures exhibit properties of softness and slenderness, from biological materials like cilia [38] and cellular membranes [39], to strands of DNA [40] and hydrogels [41]. In this instance ‘soft’ refers to the constitutive properties of the material, like that of rubber or biological tissue, and ‘slender’ refers to its geometrical thinness.

- Also include references to research that have been done using cosserat theory.

Explain that by three-dimensional rod, you mean a continuum body

Something more should be said here, prefacing the geometricisation programme. Below are some notes and random scribbles.

We provide a general framework with which to derive the - kinodynamical - kinematic and dynamical - equations of motion of continuum system with both internal and external degrees of freedom. For instance, the purview of continuum mechanics is the dynamics of sub-manifolds of \mathbb{E}^3 . Continuum mechanics thus considers systems where the external configuration space is \mathbb{E}^3 , but there are no internal degrees of freedom. Our framework can generalise these dynamics to settings where the point-continua have internal degrees of freedom, such as orientation or spin.

Kinodynamics, which is a term we borrowed from the field of robotics [157, 158], refers to the combined

Extending the continuum mechanics analogy, we can consider a system that is a sub-manifold $M_{\text{ext}} \subset \mathbb{X}$ of some homogeneous space \mathbb{X} , corresponding to the external degrees of freedom of the system. Each point-continua M_{ext} can also have an internal configuration, which can take values in some Lie group H . When we can write $\mathbb{X} = G/H$, then the configuration of the whole system described as $M \subset G$, which includes both external and internal degrees of freedom. We call this class of system *generalised Cosserat media*. An example is that of a Cosserat rod, which is a 1-dimensional rod embedded in 3-dimensional Euclidean space, where at each material point along the rod there is attached an orthonormal frame $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$. We thus have $\mathbb{X} = \mathbb{E}^3 = SE(3)/SO(3)$, and $H = SO(3)$, and $G = SE(3)$. We also consider the case when the total kinematic configuration space is not a Lie group.

We consider the dynamics of systems that can be described as sub-manifolds of a Lie group G . In the case where there is a natural division G/H and H , corresponding to external and internal degrees of freedom respectively, we call this class of systems *generalised Cosserat media*.

By kinematics we mean that equations of motion that upon integration ensures that the system remains within its own configuration space. By dynamics we mean that the dynamical equations of motion are, in the conservative case, Lagrangian equations of motion for a given Lagrangian. We also generalise the dynamics to the non-conservative case.

References:

soft robotics [159, 160, 161, 162, 163, 164, 165, 166, 167]

cell membranes [39, 168]

biological tissue [169, 170]

soft matter:

active filaments [171, 172, 173, 174]

filaments [175, 176, 177, 178]

Many applicaitons in engineering and continuum mechanics: soils [179], polycrystalline [180] and composite materials [181], granular and powder-like materials [182, 183, 184], masonries [185], cellular [186] or porous media and foams [187], bones [188], liquid crystals [189, 190], as well as electromagnetic and ferromagnetic media [191, 192, 193]

4.1 Classical Cosserat theory

Classical continuum mechanics study elastic materials as manifolds $\mathcal{M} \subset \mathbb{E}^3$ of point-continua $\mathbf{x} \in \mathcal{M}$, where $\mathbb{E}^3 \cong \mathbb{R}^3$ is 3-dimensional Euclidean space. Points in the material have three translational degrees of freedom, and the elastic response to displacements away from the rest configuration is determined by the symmetric Cauchy stress 2-tensor $\boldsymbol{\sigma}$. The

momentum transport within the continuum is given by the *Cauchy momentum equation*

$$\rho \frac{D\mathbf{v}}{Dt} = \nabla \cdot \boldsymbol{\sigma} + \mathbf{f} \quad (4.1)$$

where ρ is the mass density, \mathbf{v} is the flow velocity field, \mathbf{f} are body forces per unit volume (e.g. gravity $\mathbf{f} = \rho\mathbf{g}$), and D/Dt the material derivative. Eq. (4.1) is a non-linear PDE in three spatial dimensions and time. In the treatment of a given system, these equations of motion must thus be satisfied at each material point $\mathbf{x} \in \mathcal{M}$ as well as obey necessary conditions at the boundaries $\partial\mathcal{M}$ of the body.

However, many systems have geometric properties that make them amenable to simplified treatments. Slender systems, for instance, are ‘thin’ in one or more spatial dimensions. This often makes it feasible to model slender shells in two spatial dimensions and time, and slender rods in one spatial dimension and time. In such cases the continuum configuration of the bulk can then be suitably substituted with other, model-specific, internal degrees of freedom.

The suite of models proposed by the Cosserat brothers [1] in 1909 are one of the more prominent approaches by which one can study slender materials. Their approach is that of an extended continuum theory that, in addition to the manifold of point-continua \mathcal{M} , included a set of *directors* (three-dimensional vectors) at each point $\mathbf{x} \in \mathcal{M}$. We thus refer to these systems as *directed media*, as opposed to the undirected media of classical continuum mechanics. The physical interpretation of the directors are a matter of constitutive modelling. For three-dimensional continuum bodies, the directors could represent polar continua which has been used to study the properties of asymmetric Cauchy stress tensors, as in the case of liquid crystals [194].

When \mathcal{M} is a lower-dimensional sub-manifold of \mathbb{R}^3 , the polar continua would then often represent the substitutive internal degrees of freedom of the neglected bulk continuum. In the elastic theory of shells, a single director can model the material fibre running across its thickness, and for rods two directors are used to model the material fibres through the cross-section. We now continue to define the latter in detail.

A Cosserat rod can be defined as a curve in Euclidean space $\mathbf{r}(u) \in \mathbb{E}^3$, $u \in [0, L_0]$, known as the *center-line* and where L_0 is (in a dynamical setting) the *rest-length* of the rod, and an orthogonal triad $E(u) = (\mathbf{e}_1(u) \ \mathbf{e}_2(u) \ \mathbf{e}_3(u))$. The *material frame* $E(u)$ represents the, in general ellipsoidal, cross-section of the rod at each *material point* u along the center-line. The vectors $\mathbf{e}_2(u)$ and $\mathbf{e}_3(u)$ are the aforementioned directors of the Cosserat rod, which can vary in length and represent the semi-minor and semi-major axes of the cross-section at u . The vector $\mathbf{e}_1(u)$ is normal to the cross-section at u and is defined as $\mathbf{e}_1(u) = (\mathbf{e}_2(u) \times \mathbf{e}_3(u)) / |\mathbf{e}_2(u) \times \mathbf{e}_3(u)|$, where \times is the 3-dimensional cross-product and $|\cdot|$ is the standard norm in Euclidean space.

The deformations and rotations of the center-line and material frame comprise the full

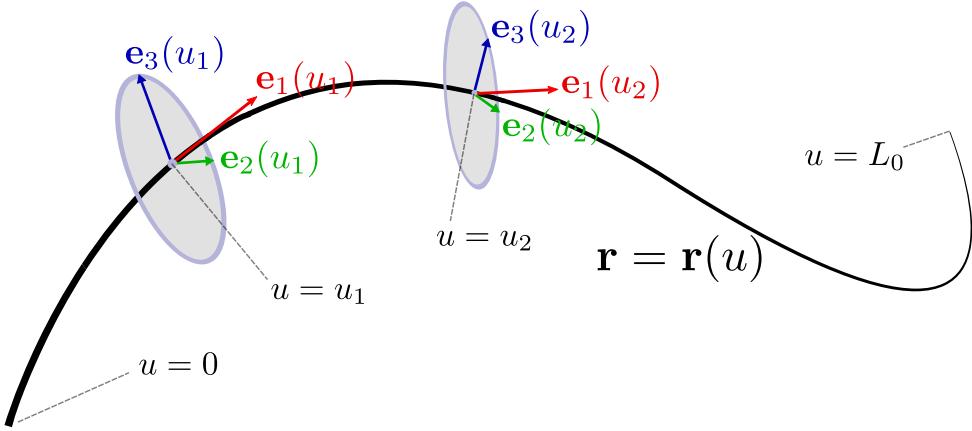


Figure 4.1: Illustration of the kinematics degrees of freedom of the Cosserat rod. The black line is the *center-line* $\mathbf{r} = \mathbf{r}(u)$ where $u \in [0, L_0]$ is the *material coordinate* along the length of the rod, where L_0 is a positive real number. When specifying constitutive dynamics, L_0 often becomes the *rest-length* of the rod. Two cross-sections at $u = u_1$ and $u = u_2$ are shown with the material frame attached $E = (\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3)$, which are the red, green and blue arrows respectively, of which the latter two are the *directors* of the rod. Note that having two directors, as opposed to a single director normal to the cross-section, allows for a twisting degree of freedom as can be seen in Fig. 4.2d.

kinematic degrees of freedom of the Cosserat rod. Often in applications the cross-section is approximated to be of constant shape along the rod, in which case we restrict the directors to be inextensible and orthogonal, and thus $E(u)$ is then an orthonormal triad. See Fig. 4.1 for an illustration of a Cosserat rod with circular cross-section. In the literature this class of Cosserat rod is known as a *special Cosserat rod* [181, 195, 196]. Henceforth, unless stated otherwise, by Cosserat rod we mean a special Cosserat rod.

The kinematic degrees of freedom of the rod are thus: smooth translations of the curve \mathbf{r} and smooth rotations the material frame-field E . In other words, the center-line can bend, and the cross-section can shear and twist around the center-line and extend tangentially across its length, as is illustrated in Fig. 4.1. The extension of the rod can be captured by the scalar $h(u) = |\frac{\partial \mathbf{r}}{\partial u}|$, which can be seen as the square root of the metric on the center-line induced by the Euclidean metric on \mathbb{R}^3 . h relates the material coordinate u to the arc-length coordinate s as

$$ds = h(u)du \quad (4.2)$$

such that $|\frac{\partial \mathbf{r}}{\partial s}| = 1$ for all $s \in [0, L]$ where

$$L[h(u)] = \int_0^{L_0} h(u)du \quad (4.3)$$

is the total arc-length of the center-line. Thus any point u for which $h(u) = 1$ is not suffering an extension. Shear and twist deformations can be distinguished by noting that the former denotes the orientation \mathbf{e}_1 of the cross-section deviating from being parallel

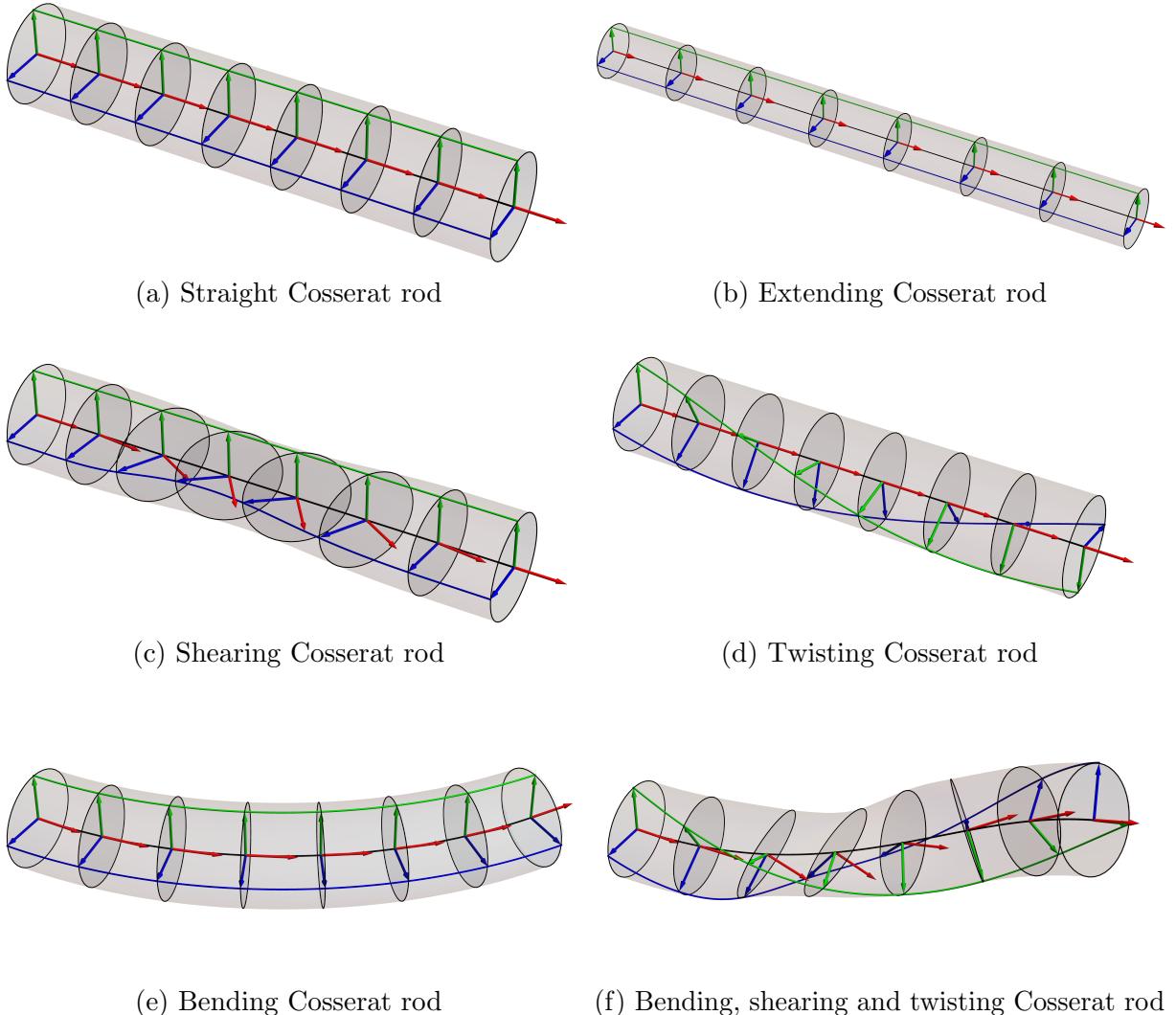


Figure 4.2: Examples of deformations of the Cosserat rod. The transparent gray body is the bulk of the Cosserat rod, with the center-line (black line) running through its radial center. The material frame and cross-section are shown at intermittent points along the center-line, along with the surface fibres traced out by the directors e_2 and e_3 , green and blue respectively. The tubular radii of the rods depicted were exaggerated in size for illustrative purposes. (a) A straight Cosserat rod suffering no deformation. (b-f) Examples of extension, shearing, twisting and bending deformations.

to the center-line, whilst the latter denotes rotations of the material frame around $\mathbf{e}_1(u)$. Henceforth we will also distinguish the length elements ds and du as the *length element* and *material length element* respectively.

We now add time to the picture, and consider the motion of the Cosserat rod as the result of arbitrary translational and angular velocity fields. Let the center-line $\mathbf{r}(t, u)$ and the material frame $\mathbf{e}_i(t, u)$ be functions of time, then the temporal evolution of the rod is

$$\dot{\mathbf{r}} = \mathbf{V} \quad (4.4a)$$

$$\dot{\mathbf{e}}_i = \boldsymbol{\Omega} \times \mathbf{e}_i \quad (4.4b)$$

from initial boundary conditions at $t = 0$, and where $\mathbf{V} = \mathbf{V}(t, u)$ and $\boldsymbol{\Omega} = \boldsymbol{\Omega}(t, u)$ are arbitrary translation and angular velocities.

Dynamical equations of motion for the Cosserat rod are found by imposing balance laws on the momentum and moments of the polar continua. For undirected media, the conservation of mass and linear momentum balance are used to determine the dynamics, given constitutive and body forces. For directed media, in addition to the above, further conservation laws must be imposed to establish the director dynamics.

For a Cosserat rod with mass density ρ_0^V and cross-sectional area A in its reference configuration, the linear momentum of the rod is $\mathbf{P} = \rho_0^V A \dot{\mathbf{V}}$, as is the case in classical continuum mechanics. We also introduce the director angular momentum $\mathbf{L} = I\boldsymbol{\Omega}$, where $I \in \mathbb{R}^{3 \times 3}$ is a moment of inertia matrix for the cross-section. As will be shown in Sec. 5, imposing the conservation of the linear momentum of the rod and the director angular momentum leads, as was first derived in [1], to

$$\dot{\mathbf{P}} = \mathbf{F}' + \mathbf{f} \quad (4.5a)$$

$$\dot{\mathbf{L}} = \mathbf{M}' + \mathbf{r}' \times \mathbf{F} + \mathbf{m}, \quad (4.5b)$$

$$\mathbf{F} = 0, \text{ at } u = 0 \text{ and } u = L_0 \quad (4.5c)$$

$$\mathbf{M} = 0, \text{ at } u = 0 \text{ and } u = L_0 \quad (4.5d)$$

where \mathbf{F} are the constitutive forces acting on the rod, \mathbf{f} the body forces per unit material length, \mathbf{M} the constitutive director moments and \mathbf{m} the body moment per unit material length. Eq. 4.5a is in a form familiar to classical continuum mechanics, which can be seen by comparing it to Eq. 4.1, whilst Eq. 4.5b is particular to the setting of directed media. Equation 4.5 and Eq. 4.4 together form a closed set of first-order equations in time and space for the kinodynamics of a Cosserat rod.

We note here that up until this point we have considered an open Cosserat rod where $\mathbf{r}(0, t) \neq \mathbf{r}(L_0, t)$. The dynamics of a closed Cosserat rod, for which the center-line and frame are periodic functions of u , are identical with the exception of the omission of the

boundary conditions Eq. 4.5c and Eq. 4.5d.

Equation 4.5 can be derived directly from the balance laws of classical continuum mechanics, as shown in [196, 197], under the kinematic assumption that the cross-sections traced out by the directors correspond to the bulk of a three-dimensional rod of undirected point-continua. For illustrative purposes this kinematic assumption deserves further elaboration. In precise mathematical language, let $\mathcal{M} \subset \mathbb{E}$ and let $\mathbf{x} : D \rightarrow \mathcal{M}$ be the *material coordinate* function from the domain

$$D = \{(X_1, X_2, X_3) : X_1^2 + X_2^2 = R^2 \text{ and } X_3 \in [0, L_0]\} \quad (4.6)$$

where R is a given fixed tubular radius of the Cosserat rod. Given a Cosserat rod configuration (\mathbf{r}, E) , we define \mathcal{M} via the material coordinate function as

$$\mathbf{x}(\mathbf{X}) = \mathbf{r}(X_3) + X_j \mathbf{e}_j, \quad j = 2, 3 \quad (4.7)$$

such that \mathcal{M} is the image of \mathbf{x} . Here the material coordinate X_3 corresponds to the coordinate u . We see how the Cosserat rod can be viewed as the result of a coarse-graining procedure from the full three-dimensional continuum setting, replacing the cross section at each u with two directors, thus reducing the spatial coordinates of the system from three to one.

We now conclude this section by making some remarks, prefacing the discussions in subsequent chapters, on the geometric properties of the Cosserat rod. The kinematic equations of motion Eq. 4.4a and Eq. 4.4b shows explicitly that the rod moves according to infinitesimal translations $\mathbf{V} dt$ and rotations $\boldsymbol{\Omega} \times \mathbf{e}_i$ respectively. This entails that we can identify the kinematic structure of the Cosserat rod with the *Lie group* of Euclidean transformations of translations and rotations $SE(3)$ [198, 199]. In particular the rod itself can be parametrised in terms of sub-manifolds of $SE(3)$ and, as will be the main subject of Sec. 5, from which geometricised kinodynamical equations of motion can be derived programmatically. In the following section some mathematical foundations will be introduced, necessary for the subsequent chapters of this part of the thesis.

4.2 Mathematical preliminaries

As will be further discussed in Ch. 5, directed media can be seen as either sub-manifolds of *homogeneous spaces* or sub-manifolds of *Lie groups*. Through this lens, a fully geometricised and non-coordinate form of the equations of motion of such systems can be derived. This section serves primarily to establish the mathematical foundation and rigour of the programme, and therefore has a level of mathematical abstraction higher than that of the subsequent chapters. The reader not interested in these details may proceed to Ch. 5,

and return to this section intermittently to fill gaps in notation and conceptual knowledge. For a fuller treatment of the concepts introduced in this section the reader can consult the following references for further exposition [42, 200, 201, 202].

4.2.1 Differential geometry

This will be an abbreviated exposition of differential geometry. See [203] for more details. Let M be a set. A *chart* on M is a pair (U, ϕ) where $U \subset M$ and ϕ is a bijective map from U to a subset $\phi(U) \subset \mathbb{R}^n$. The *atlas* \mathcal{A} is the set of all charts on M . If for any two charts $(U_1, \phi_1), (U_2, \phi_2) \in \mathcal{A}$ we have that the map $\phi_2 \circ \phi_1^{-1} : \phi_1(U_1 \cap U_2) \rightarrow \phi_2(U_1 \cap U_2)$ is a smooth function, then M is called an n -dimensional *smooth manifold*. Henceforth by *manifold*, we will mean a smooth manifold.

A function $f : M \rightarrow \mathbb{R}$ is smooth if $f \circ \phi^{-1}$ is a smooth function for any chart $(U, \phi) \in \mathcal{A}$. A curve $\gamma : [0, 1] \rightarrow M$ in M is a *smooth curve* if $\phi \circ \gamma$ is smooth for any chart $(U, \phi) \in \mathcal{A}$ along the intersection $\gamma([0, 1]) \cap U$.

Given a map $\psi : M \rightarrow N$, where M and N are m - and n -dimensional manifolds respectively. Let $f : N \rightarrow \mathbb{R}^n$ be a smooth function. Then the *pull-back* $\psi^* f : M \rightarrow \mathbb{R}^m$ of f under ψ is defined as

$$\psi^* f = f \circ \psi. \quad (4.8)$$

Let $f : M \rightarrow \mathbb{R}$ be a smooth function and let $\gamma : [0, 1] \rightarrow M$ be a smooth curve, and let $\gamma(\alpha_0) = p$, where $\alpha_0 \in (0, 1)$. The *tangent vector* to γ at p is defined as

$$X_p(f) = \left. \frac{d}{d\alpha} f(\gamma(\alpha)) \right|_{\alpha=\alpha_0} \quad (4.9)$$

which is a linear map from the space of smooth functions on M to \mathbb{R} . The *tangent space* $T_p M$ is the set of all such linear maps at a point $p \in M$, which forms an n -dimensional vector space [203]. The set of all tangent spaces

$$TM = \{(p, X_p) : p \in M, X_p \in T_p M\}. \quad (4.10)$$

is known as the *tangent bundle*.

A *vector field* $X : M \rightarrow TM$ is a map that assigns a tangent vector $X(p) \in T_p M$ at each point $p \in M$. For any smooth function $f : M \rightarrow \mathbb{R}$ and a vector field X , we define $Xf : M \rightarrow \mathbb{R}$ as the function defined by

$$p \mapsto X_p(f). \quad (4.11)$$

Therefore Xf should be seen as a differentiation of f . If the function Xf is smooth for all smooth functions f , then we say that X is a *smooth vector field*. We denote the space of

all smooth vector fields by $\Gamma(TM)$.

Given a chart $(U, \phi) \in \mathcal{A}$, which we write in component form as $\phi(p) = (x_1(p), x_2(p), \dots, x_n(p))$, then we can expand vector field locally on U as [203]

$$X = X^i \frac{\partial}{\partial x^i} \quad (4.12)$$

where $X^i : M \rightarrow \mathbb{R}$ are smooth functions we call the *coefficients* of X .

4.2.2 Exterior calculus

In calculus, the differential of a scalar function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ is often written as

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz. \quad (4.13)$$

This can be generalised for smooth manifolds. Let M be a smooth d -dimensional manifold, $p \in M$ a point on the manifold, $X \in \Gamma(TM)$ a vector field on M and $f \in C^\infty(M)$ be a smooth function on M . We define the map $df : TM \rightarrow \mathbb{R}$ as

$$(df(X))(p) = X_p(f). \quad (4.14)$$

Here df is an example of a *scalar-valued 1-form*. Analogously, scalar functions $f \in C^\infty(M)$ are known as *scalar-valued 0-forms*. Scalar-valued p -form are linear maps $\prod_{i=1}^p TM \rightarrow \mathbb{R}$, where \prod signifies the repeated Cartesian product. We will therefore often use the operator d is the *exterior derivative*, which in general maps p -forms to $(p+1)$ -forms. Any $(p+1)$ -form that can be written as the exterior derivative of a p -form is referred to as *exact*. Conversely, any p -form ϕ that satisfies $d\phi = 0$ is *closed*. Locally, closed p -forms are always exact, which is known as the *Poincaré lemma*. As 1-forms ϕ are maps $\phi : TM \rightarrow \mathbb{R}$, we will often refer to them as *covector fields*. We write $\phi \in \Gamma(T^*M)$ and $\phi(p) \in T^*_p M$.

Let $x^i : U \rightarrow \mathbb{R}$, $i = 1, \dots, d$ be coordinate functions for some neighbourhood $U \subset M$. Then df can then be locally expressed in U as

$$df = \frac{\partial f}{\partial x_i} dx^i \quad (4.15)$$

such that $df(X) = \frac{\partial f}{\partial x^i} dx^i(X) = \frac{\partial f}{\partial x^i} X^i$, where $X \in \Gamma(TM)$. Not all 1-forms are closed, and can thus not be written in the form of Eq. 4.15, but they can always be expanded in a coordinate basis as $\phi = a_i dx^i$, where $a_i \in C^\infty(M)$, $i = 1, \dots, d$ are the coefficients of ϕ in this basis.

The *symmetric product* of a p -form x and q -form y is written as $x \otimes y$. For $p = q = 1$,

we have

$$(x \otimes y)(X, Y) = \phi(X)\psi(Y) \quad (4.16)$$

where $X, Y \in \Gamma(TM)$ are two vector fields. For a general 1-form, its exterior derivative can be expressed locally as

$$d\phi = da_i \wedge dx^i. \quad (4.17)$$

where \wedge is the *wedge product*. Let x and y be a p -form and a q -form respectively, then $x \wedge y$ is a $(p+q)$ -form defined as

$$x \wedge y = x \otimes y - y \otimes x \quad (4.18)$$

and satisfies

$$y \wedge x = (-1)^{pq}x \wedge y. \quad (4.19)$$

and

$$d(x \wedge y) = dx \wedge y + (-1)^p x \wedge dy. \quad (4.20)$$

The wedge product of a sequence of forms can be written as

$$\bigwedge_{i=1}^n z^i = dz^1 \wedge dz^2 \wedge \cdots \wedge dz^n \quad (4.21)$$

where each z^i is a p^i -form. Due to the anti-symmetry of the wedge products of 1-forms, we have that $df \wedge df = 0$ for any $f \in C^\infty(M)$. If x and y are two 1-forms, then the product $x \wedge y$ is a mapping $TM \times TM \rightarrow \mathbb{R}$, and can be evaluated on two vector fields $X, Y \in \Gamma(M)$ as

$$(x \wedge y)(X, Y) = x(X)y(Y) - x(Y)y(X). \quad (4.22)$$

We will now derive a coordinate-free expression for the exterior derivative of a 1-form. Let us first define the *Lie bracket of vector fields* $[\cdot, \cdot]$, which is

$$[X, Y](f) = X(Y(f)) - Y(X(f)) \quad (4.23)$$

and can be expressed in component form it as [203]

$$[X, Y](f) = (X^j \partial_j Y^i - Y^j \partial_j X^i) \partial_i f. \quad (4.24)$$

Now, let $\phi = a_i dx^i$, then

$$\begin{aligned}
d\phi(X, Y) &= da_i(X)Y^i - da_i(Y)X^i \\
&= \frac{\partial a_i}{\partial x^k} X^k Y^i - \frac{\partial a_i}{\partial x^k} Y^k X^i \\
&= \left(X^k \frac{\partial}{\partial x^k} (a_i Y^i) - a_i X^k \frac{\partial Y^i}{\partial x_k} \right) - \left(Y^k \frac{\partial}{\partial x^k} (a_i X^i) - a_i Y^k \frac{\partial X^i}{\partial x_k} \right) \\
&= X^k \frac{\partial}{\partial x^k} (a_i Y^i) - Y^k \frac{\partial}{\partial x^k} (a_i X^i) - a_i \left(X^k \frac{\partial Y^i}{\partial x_k} - Y^k \frac{\partial X^i}{\partial x_k} \right) \\
&= X(\phi(Y)) - Y(\phi(X)) - \phi([X, Y]).
\end{aligned} \tag{4.25}$$

Intuitively, a 1-form can be seen as measuring an infinitesimal oriented length. A 2-form can be seen as measuring an infinitesimal oriented area. A p -form measures an infinitesimal oriented p -dimensional volume. There is therefore a natural notion of integrals of p -forms. In general, a p -form ϕ defined on a manifold M can be integrated on a p -dimensional sub-manifold of M .

If M is d -dimensional, then the integration of d -forms and 1-forms coincides with the usual notion of integration in multi-variate calculus. Let $U \subset M$ with coordinates $x^i : U \rightarrow \mathbb{R}^d$, and let $\gamma \subset U$ be a 1-dimensional sub-manifold of U , and let $\phi = h_i dx^i$ be a general 1-form and $\psi = g \bigwedge_{i=1}^d dx^i$ a general d -form then

$$\int_{\gamma} \phi = \int_{\gamma} h_i dx^i \tag{4.26a}$$

$$\int_U \psi = \int_U g \bigwedge_{i=1}^d dx^i = \int_U g \, dx^1 dx^2 \dots dx^d \tag{4.26b}$$

where the right-most term in the equalities are the standard multi-variate integrals. For a d -dimensional manifold, a d -form like ψ is called a *volume form*. If $V \subset M$ is a $(p-1)$ -dimensional sub-manifold and ϕ is a p -form, then it can be shown that

$$\int_V d\phi = \int_{\partial V} \phi \tag{4.27}$$

where ∂V signifies the boundary of V . Eq. 4.27 is called *Stokes' theorem*.

Consider a 1-form expressed locally as $\phi = a_i dx^i$. Under change of coordinates, it transforms as

$$a_i dx^i = a_i \frac{\partial x^i}{\partial \tilde{x}^i} d\tilde{x}^i. \tag{4.28}$$

From Eq. 4.28, it can be shown that d -forms transform as

$$f \bigwedge_{i=1}^d dx^i = \left(\det \left[\frac{\partial \mathbf{x}}{\partial \tilde{\mathbf{x}}} \right] f \right) \bigwedge_{i=1}^d d\tilde{x}^i \tag{4.29}$$

where $\frac{\partial \mathbf{x}}{\partial \tilde{\mathbf{x}}}$ is the Jacobian matrix of the coordinate transformation.

4.2.3 Lie groups

Here we will define and state relevant results on *Lie group* theory, which will be used extensively throughout this text. For many of the systems we consider, we will identify their configurations with elements of Lie groups. Here we will only give a brief overview of theory. See [200, 204, 205] for a more detailed treatment.

A d -dimensional *Lie group* is a set G that is both a group and a smooth manifold of dimensions d , where the multiplication map $G \times G \rightarrow G$

$$(g, h) \mapsto gh \in G \quad (4.30)$$

and inverse $G \rightarrow G$

$$g \mapsto g^{-1} \quad (4.31)$$

are smooth maps.

For all Lie groups G under consideration in this text the group elements in G can be represented as invertible matrices. In other words, there is a map $\Pi : G \rightarrow GL(V)$, where $GL(V)$ is the Lie group of linear operators with non-zero determinant acting on some n -dimensional vector space V . Such a map is called a *representation* of G . For all Lie groups there is a canonical *fundamental representation* $g : G \rightarrow GL(\mathbb{R}^n)$, and we will often identify the image $g(G) = \tilde{G}$ with the Lie group itself. Henceforth, unless explicitly stated to be otherwise, we will use the shorthand $g \in G$, identifying the Lie group elements in their fundamental representation with the Lie group elements themselves.

For each element $g \in G$, the *left multiplication* map $L_g : G \rightarrow G$ is defined as $L_g h = gh$ where $h \in G$. Similarly, the *right multiplication* map $R_g : G \rightarrow G$ is defined as $R_g h = hg$. For maps between smooth manifolds $\Psi : M \rightarrow N$, we define its *derivative* at $p \in M$ as the mapping $D\Psi_p : T_p M \rightarrow T_{\Phi(p)} N$ given by the formula

$$D\Psi_p(v)(f) = v(f \circ \Psi) \quad (4.32)$$

for $v \in T_p M$ and $f \in C^\infty(N)$. In particular, the derivative of the left multiplication at $h \in G$ is a mapping $(DL_g)_h : T_h G \rightarrow T_{gh} G$, which can be shown to be equal to

$$(DL_g)_h(X) = gX \quad (4.33)$$

where $X \in T_h G$. For any Lie group G there is an associated *Lie algebra* \mathfrak{g} , which is the tangent space $T_e G$ at the identity, where $e \in G$ is the identity element. \mathfrak{g} is thus a vector space of the same dimension d as G . For any vector $W \in \mathfrak{g}$, we can define a *left-invariant*

vector field \tilde{W} defined on each $g \in G$ as

$$\tilde{W}_g = (DL_g)_e(W). \quad (4.34)$$

Now let E_i , $i = 1, \dots, d$ be a basis for \mathfrak{g} , then it can be shown that the corresponding left-invariant vector-fields \tilde{E}_i form a global basis for the tangent bundle TG . This implies that the tangent bundle of Lie groups G are isomorphic to the Cartesian product $TG \cong G \times \mathfrak{g}$. Furthermore, it is notable that the global basis is constructed without specifying coordinate functions on the manifold.

The *exponential map* $\exp : \mathfrak{g} \rightarrow G$ relates Lie algebra elements to corresponding Lie group elements, and for matrix Lie groups this is explicitly given by the matrix exponential. Intuitively this mapping can be understood by considering the curve $\gamma(t) = \exp(tV) \in G$, where $V \in \mathfrak{g}$. Differentiating the curve gives us

$$\begin{aligned} \frac{d}{dt}\gamma(t) &= \exp(tW)W \\ &= (DL_{\exp(tW)})_e(W) = \tilde{W}_g \end{aligned} \quad (4.35)$$

The curve γ is thus a flow-line of the left-invariant vector field \tilde{W} .

The Lie algebra \mathfrak{g} also has a product structure known as the *Lie bracket* $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$ given by

$$[X, Y] = XY - YX \quad (4.36)$$

for $X, Y \in \mathfrak{g}$. To see its relation to the Lie group, we introduce the *adjoint action* of a Lie group on its Lie algebra $\text{Ad} : G \times \mathfrak{g} \rightarrow \mathfrak{g}$ given by

$$\text{Ad}_g Y = gYg^{-1} \quad (4.37)$$

for each $g \in G$, Ad_g is thus an automorphism of the Lie algebra. Let $g(t) = \exp(tX)$, with Taylor expansion $g(t) = I + tX + O(t^2)$. The infinitesimal action of the automorphism Ad_g can then be found by expanding it as

$$\text{Ad}_g Y = Y + t[X, Y] + O(t^2). \quad (4.38)$$

This motivates the definition of the adjoint action of the Lie algebra *on itself* $\text{ad} : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$ given by

$$\text{ad}_X Y = [X, Y]. \quad (4.39)$$

Let E_a , $a = 1, \dots, d$ be a basis for Lie algebra \mathfrak{g} , then the Lie bracket can also be written

in terms of its *structure constants* $f_{bc}^a \in \mathbb{R}$

$$[E_a, E_b] = f_{ab}^c E_c, \quad a, b, c = 1, \dots, d. \quad (4.40)$$

The Lie algebra can be *defined* as a vector space \mathfrak{g} equipped with Lie bracket, where the latter is fully specified by the structure constants. If G is connected, then G is exactly equal to the image of \mathfrak{g} under the exponential map ¹. This is known as the *Lie algebra-Lie group correspondence*. We thus see that the structure constants encapsulates the full geometry of the Lie group.

For a d -dimensional Lie group G , with Lie algebra \mathfrak{g} , a linear map $\prod_{i=1}^p TG \rightarrow \mathfrak{g}$ is a *Lie algebra-valued p-form*. A general Lie algebra-valued 1-form $X : TG \rightarrow \mathfrak{g}$ can be written as

$$X = a_i^a E_a dx^i \quad (4.41)$$

where E_a , $a = 1, \dots, d$ is a basis for \mathfrak{g} , $x^i \in C^\infty(G)$, $i = 1, \dots, d$ are global coordinate functions on G , and $a_i^a \in C^\infty(G)$, $i, a = 1, \dots, d$ are the coefficients of X in this basis. The exterior derivative of a Lie-algebra valued form can be computed by simply applying it directly on the scalars in the expression. For example, the exterior derivative of a Lie algebra-valued 1-form is

$$dX = E_a d(a_i^a dx^i). \quad (4.42)$$

Similarly, let $Y = b_j^a E_a dx^j$, then the wedge product of two 1-forms can be computed as

$$X \wedge Y = E_a E_b a_i^a b_j^b dx^i \wedge dx^j \quad (4.43)$$

where $E_a E_b$ denotes ordinary matrix multiplication. Note that the wedge product of two Lie algebra-valued forms is not necessarily Lie algebra-valued.

As with Lie group elements, Lie algebra elements can be treated as matrices. For any Lie group representation $\Pi : G \rightarrow GL(V)$, there is a corresponding representation $\rho : \mathfrak{g} \rightarrow \mathfrak{gl}(V)$ of the Lie algebra, that are related as $\exp(\rho(W)) \in \Pi(G)$ for any $W \in \mathfrak{g}$. As with Lie groups, there is a canonical fundamental representation of the Lie algebra as linear maps $\mathfrak{gl}(\mathbb{R}^n)$, and we identify this representation with \mathfrak{g} itself.

Given the fundamental representation of \mathfrak{g} , the *dual* Lie algebra \mathfrak{g}^* is represented as the set of matrices [205]

$$\mathfrak{g}^* = \{Y^T : Y \in \mathfrak{g}\}. \quad (4.44)$$

Mathematically, \mathfrak{g}^* should be seen as the set of linear maps on \mathfrak{g} , via an inner product $\langle \cdot, \cdot \rangle : \mathfrak{g}^* \times \mathfrak{g} \rightarrow \mathbb{R}$, which we will define in the main text. We will not delve any deeper into the concept of the dual Lie algebra here, but we will instead let the exposition unfold

¹If G is not connected, $\exp(\mathfrak{g})$ is equal to the sub-group of G that is connected to the identity element $e \in G$.

within the main text itself.

The following are some examples of Lie groups:

- The general linear group $GL(V)$ is the set of linear maps with non-zero determinant acting on a vector space V .
- The positive real numbers \mathbb{R}^+ equipped with multiplication forms an *abelian* Lie group, where all elements commute with each other.
- The orthogonal group $O(n)$ which is the space of orthogonal $n \times n$ matrices. Its Lie algebra $\mathfrak{o}(n)$ comprises the space of anti-symmetric $n \times n$ matrices.
- The special orthogonal group $SO(n)$ is the set of matrices $R \in SO(n)$ for which $\det R = 1$. Its Lie algebra $\mathfrak{so}(n)$ is equal to $\mathfrak{o}(n)$.
- The translation group $T(V)$ is the Lie group of translations on a vector space V . We denote $T(n)$ as the group of translation on n -dimensional Euclidean space.
- The special Euclidean group $SE(n)$ is the group of translations and rotations in n -dimensional space, and can be written as the semi-direct product $SE(n) = T(n) \rtimes SO(n)$.

4.2.4 Homogeneous spaces

Here we give a brief overview of the notion and definition of a homogeneous space. Intuitively, a homogeneous space \mathbb{X} is a set that is a space that ‘looks the same’ everywhere. For example, spheres are self-similar under rotation, Euclidean space is self-similar upon translation and rotation, both of which are examples of homogeneous spaces. See [200, 204] for a more detailed treatment.

Let \mathbb{X} be a set and G a Lie group. An *action* $\phi : G \times X \rightarrow X$ is a map that satisfies $\phi(e, x) = x$ and $\phi(g, \phi(h, x)) = \phi(gh, x)$ for all $g, h \in G$ and $x \in \mathbb{X}$. We abbreviate the group action as $\phi(g, x) = gx$. We say that G acts *transitively* on \mathbb{X} if the action is such that for any pair of elements $x, y \in \mathbb{X}$, there exists a group element $g \in G$ such that $gx = y$. If a set \mathbb{X} admits a transitive action of a Lie group G , then \mathbb{X} is known as a *homogeneous space*, and G is called the *symmetry group* of \mathbb{X} .

All homogeneous spaces can be expressed as the quotient of a Lie group G and a Lie subgroup $H \subset G$ [206]. To define this notion, we first define *left coset*, which is the set

$$gH = \{gh : h \in G\}. \quad (4.45)$$

Then the *quotient space* G/H is the set

$$G/H = \{gH : g \in G\}. \quad (4.46)$$

Common examples of homogeneous spaces are Euclidean space $\mathbb{E}^d \cong SE(d)/O(d)$ and the d -sphere $S^n \cong SO(n)/SO(n-1)$.

Given a homogeneous space G/H , there is a natural projection map $\pi : G \rightarrow G/H$ defined as

$$\pi(g) = gH. \quad (4.47)$$

The triple $(G, G/H, \pi)$ is known as a *principal bundle*, where G/H is called the *base space* and G the *total space*.

4.2.5 The Maurer-Cartan form

We have seen that the Lie group G can be fully reconstructed using the structure constants f_{ab}^c of its Lie algebra \mathfrak{g} . The same geometric information is also contained in its left-invariant vector fields Eq. 4.34², which can be encapsulated using a Lie algebra-valued 1-form called the *Maurer-Cartan form* $\omega : TG \rightarrow \mathfrak{g}$, which is defined as

$$\omega(v) = (DL_{g^{-1}})_g(v) \quad (4.48)$$

for any $v \in T_g G$. The Maurer-Cartan form thus maps the tangent spaces at any point g to the tangent space at the identity \mathfrak{g} .

Let $X, Y \in \mathfrak{g}$ and let $v_X = (DL_g)_e(X)$ and $v_Y = (DL_g)_e(Y)$ be their corresponding left-invariant vector fields. Then the Maurer-Cartan form satisfies

$$\omega([v_X, v_Y]) = [X, Y] \quad (4.49)$$

where the argument of ω on the left-hand side is the Lie bracket of vector fields. Eq. 4.49 can be derived by noting that for any diffeomorphism $\Psi : M \rightarrow N$, we have that $D\Psi([v, w]) = [D\Psi(v), D\Psi(w)]$ where $v, w \in \Gamma(TM)$. By recovering the Lie bracket of the Lie algebra from the Maurer-Cartan form, we thus see that it encapsulates all geometric information about the Lie group.

From Eq. 4.25 we have that

$$d\omega(v, w) = v(\omega(w)) - w(\omega(v)) - \omega([v, w]). \quad (4.50)$$

If v and w are left-invariant, then $v(\omega(w)) = w(\omega(v)) = 0$, and we get

$$d\omega(v, w) + \omega([v, w]) = 0 \quad (4.51)$$

but as left-invariant vectors span the whole tangent bundle, this equation must also hold

²Equivalently, the right-invariant vector fields contains the same information.

for arbitrary for all vector fields. Eq. 4.51 is known as the *Maurer-Cartan equation*, and can be seen as the defining equation for ω , and will be instrumental in future chapters.

For the applications in the coming chapters, it is more convenient to work in a matrix representation of the Maurer-Cartan form, given by

$$\omega = g^{-1}dg. \quad (4.52)$$

To explain the above expression, we temporarily reintroduce the distinction between abstract Lie group elements $x \in G$, and their matrix representations $g(x) \in \tilde{G}$. Eq. 4.52 is more accurately written as $\omega_x = g(x)^{-1}Dg_x$ where $x \in G$ is here an abstract (non-matrix) Lie group element. To see the correspondence between Eq. 4.48 and Eq. 4.52, we first note that the derivatives of linear maps $\Psi : M \rightarrow N$, defined in Eq. 4.32, is equivalent to the exterior derivative of a 0-form. Thus $dg = Dg$, and $Dg_x : T_x G \rightarrow T_{g(x)}\tilde{G}$, which is computed as a matrix of 1-forms. Finally we then left-translate dg to get $g^{-1}dg : TG \rightarrow \mathfrak{g}$.

Finally, we derive Eq. 4.51 in matrix form. We take the exterior derivative of Eq. 4.52

$$\begin{aligned} d\omega &= d(g^{-1}dg) \\ &= -(g^{-1}dgg^{-1}) \wedge dg + g^{-1}d^2g \\ &= -\omega g^{-1} \wedge dg, \end{aligned} \quad (4.53)$$

to get

$$d\omega + \omega \wedge \omega = 0 \quad (4.54)$$

where we have used the fact that dg is a matrix of closed 1-forms, so that $d^2g = 0$, that the wedge product commutes with matrix multiplication, and $d(g^{-1}g) = 0$ to find an expression for dg^{-1} .

A *principal bundle*

4.2.6 Sub-manifolds of homogeneous spaces

The exposition in this section follows closely that of [200]. The Maurer-Cartan form and the Maurer-Cartan equation are particularly useful when studying sub-manifolds of homogeneous spaces. If $M \subset G/H$, let $\Phi : W \rightarrow M$ be a differentiable map, where W is some n -dimensional manifold, where \mathbb{T}^n is the n -dimensional torus, for some n . The *pullback bundle* Φ^*G of the principal bundle $\pi : G \rightarrow G/H$ is defined as

$$\Phi^*G = \{(u, g) \in W \times G : \Phi(u) = \pi(g)\}, \quad (4.55)$$

i.e. Φ^*G is a principal bundle over W , such that its fibre over $u \in W$ is equal to the fibre of G over $\Phi(u) \in G/H$. Φ^*G has the map $\hat{\Phi} : \Phi^*G \rightarrow G$

$$\hat{\Phi}(u, g) = g \quad (4.56)$$

A *lifting* $\tilde{\Phi} : W \rightarrow G$ is a map that satisfies

$$(\pi \circ \tilde{\Phi})(u) = \Phi(u), \quad \forall u \in W. \quad (4.57)$$

In subsequent chapters, Φ will be identified with the space-time kinematic configuration of a system, and the lifting $\tilde{\Phi}$ a G -valued field on W that reconstructs Φ . The homogeneous space G/H is then the kinematic configuration space of the system, and $W = (\text{time domain}) \times (\text{material domain})$ will be referred to as the *kinematic base space*. To make this analogy explicit, consider a Cosserat rod with material coordinate $u \in [0, L_0]$ and temporal coordinate $t \in [0, T]$, then $W = [0, T] \times [0, L_0]$ and $\Phi(t, u)$ is the (G/H -valued) configuration of the material point u at time t . The manifolds $[0, L_0]$ and $[0, T]$ will be called the *material base space* and the *time domain* respectively.

Similarly to how the Maurer-Cartan form can be shown to contain the geometrical information of a Lie group G , we will now also show that it can play a similar role for Φ . Let $\tilde{\Phi} : W \rightarrow G$ be a differentiable map and let $\xi := \tilde{\Phi}^*\omega$ be the pull-back of the Maurer-Cartan form onto W . As $\omega : TG \rightarrow \mathfrak{g}$, we have that $\xi : TW \rightarrow \mathfrak{g}$. To derive the matrix expression for ξ we note that for any $v \in T_u W$ and $u \in W$ we have

$$\begin{aligned} \xi(v) &= \tilde{\Phi}^*\omega(v) = \omega(D\tilde{\Phi}(v)) = (DL_{\tilde{\Phi}(u)^{-1}})_{\tilde{\Phi}(u)}(D\tilde{\Phi}(v)) \\ &= \tilde{\Phi}(u)^{-1}(D\tilde{\Phi})_u(v) \end{aligned} \quad (4.58)$$

where the definition of the Maurer-Cartan form Eq. 4.48. We can thus write

$$\begin{aligned} \xi &= \tilde{\Phi}^{-1}d\tilde{\Phi} \\ &= g^{-1}dg, \quad g \in \tilde{\Phi}(W) \end{aligned} \quad (4.59)$$

where used $D\tilde{\Phi} = d\tilde{\Phi}$. The following lemma will show that all information in Φ is encapsulated by ξ , up to global transformations $g \in G$.

Lemma 4.2.1. *Let W be an n -dimensional smooth manifold, and let $\tilde{\Phi}_1, \tilde{\Phi}_2 : W \rightarrow G$ be two differentiable maps, and $\xi_i = \tilde{\Phi}_i^*\omega$, $i = 1, 2$. Then there exists a $g \in G$ such that*

$$\tilde{\Phi}_1(u) = g\tilde{\Phi}_2(u), \quad \forall u \in U \quad (4.60)$$

if and only if $\xi_1 = \xi_2$ where ω is the Maurer-Cartan form of G .

Proof. The following is a reproduction of results in [200], but where we have relaxed the

condition that $\dim(W) = n \leq \dim(G/H)$, to allow for arbitrary n . There exists a function $f : W \rightarrow G$ such that

$$\tilde{\Phi}_1(u) = f(u)\tilde{\Phi}_2(u), \quad \forall u \in W. \quad (4.61)$$

Differentiating the above, we get

$$d\tilde{\Phi}_1 = df\tilde{\Phi}_2 + fd\tilde{\Phi}_2 \quad (4.62)$$

Thus we have

$$\begin{aligned} \xi_1 &= \tilde{\Phi}_1^*\omega = \tilde{\Phi}_1^{-1}d\tilde{\Phi}_1 \\ &= \tilde{\Phi}_1^{-1}(df\tilde{\Phi}_2 + fd\tilde{\Phi}_2) \\ &= \tilde{\Phi}_1^{-1}df\tilde{\Phi}_2 + \tilde{\Phi}_1^{-1}f\tilde{\Phi}_2\tilde{\Phi}_2^{-1}d\tilde{\Phi}_2 \\ &= \tilde{\Phi}_1^{-1}df\tilde{\Phi}_2 + \xi_2, \end{aligned} \quad (4.63)$$

from which it follows that $\xi_1 = \xi_2$ if and only if $df = 0$. \square

The same steps taken to derive Eq. 4.53 can be repeated to find that ξ satisfies

$$d\xi + \xi \wedge \xi = 0. \quad (4.64)$$

which are in this context also sometime referred to as the Maurer-Cartan equations. More precisely, they can be seen as the pull-back of the Maurer-Cartan equations on to W . The above lemma, along with Eq. 4.64, are the corner-stones of the general kinematic theory of directed continuum materials, whose kinematic configuration space is a Lie group or homogeneous space, which we will develop in the subsequent chapters. It tells us that *all geometric information in Φ is encoded in ξ* , subject to global transformations Eq. 4.60, and furthermore that ξ is invariant under said transformations.

In Cartan's *theory of moving frames*, the Maurer-Cartan form is utilised to establish the equivalence of sub-manifolds of homogeneous spaces [200, 207, 208, 209]. For a given homogeneous space G/H , d -dimensional sub-manifolds can be characterised by a set of *differential invariants*, which are scalar functions defined on the sub-manifold. These differential invariants capture the intrinsic geometry of the sub-manifold, and are invariant under transformations in H . For example, space-curves $\gamma : [0, 1] \rightarrow \mathbb{R}^3$ can be seen as sub-manifolds the homogeneous space $SE(3)/SO(3) \cong \mathbb{R}^3$. Any such curve is uniquely determined, up to global rigid-body rotations in $SO(3)$, by its *torsion* $\tau(u)$ and *curvature* $\kappa(u)$, where $u \in [0, 1]$. The equations that recover γ from the two scalar differential invariants are the celebrated *Frenet-Serret* equations. In the language of physics, such a space curve is called a *filament*, and will be treated in Ch. 5.

However, the relevance of the above results go beyond being a mathematical nicety. In simulations, working directly with Lie group-valued objects like Φ can be impractical. As all Lie groups G will in practice be represented as a sub-Lie group of $GL(\mathbb{R}^{n \times n})$, numerical

errors will in general take values in the space of n -by- n matrices. Let $h \in \mathbb{R}^{n \times n}$ represent the error accrued in a single time-step, and let $g \in G$ represent the current state of a simulation. Then it is clear that $g + h \notin G$. This is inherently due to the non-linearity of the space G . In the later chapters, we will instead use the Maurer-Cartan form to formulate the kinodynamics of sub-manifolds of G in terms of its Lie algebra \mathfrak{g} . The state of the system thus take values in a linear space, as do the errors, which ensures that in simulations the system never falls out of the correct space.

We have considered a map Φ and showed its relation to the pullback of the Maurer-Cartan form ξ . In applications we will most often have some $\xi : TW \rightarrow \mathfrak{g}$ and then reconstruct its corresponding Φ . From Eq. 4.59, we have

$$d\tilde{\Phi} = \tilde{\Phi}\xi. \quad (4.65)$$

This is a matrix ODE that can be solved numerically.

Note that for many systems the kinematic configuration space will be an entire Lie group G (e.g. a Cosserat rod), rather than a homogeneous space G/H (e.g. a filament, without a material frame). In this case we can use that $G \cong G/e$ and all of the above will still apply. Consequently, we must have $\tilde{\Phi} = \Phi$ as $G = G/H$. Conversely if H is not trivial there is an infinite dimensional space of admissible $\tilde{\Phi}$ for a given Φ . This latter fact will later be described as a *gauge freedom* in the kinematic description of the system.

4.2.7 Vector and matrix operations

We conclude the mathematical preliminaries with a list of some useful vector and matrix operations, that will be used frequently throughout this part of the thesis. In three-dimensional space, there is a natural isomorphism between 3-vectors and anti-symmetric 3×3 -matrices. This isomorphism is known as the hat map. For a column vector $\vec{v} \in \mathbb{R}^3$, the corresponding anti-symmetric matrix \hat{v} is defined as

$$\hat{v} = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix}. \quad (4.66)$$

Conversely, for a given anti-symmetric matrix $\hat{b} \in \mathbb{R}^{3 \times 3}$, we denote the inverse hat-map as \vec{v} , given by

$$\vec{b} = (b_{32} \ b_{13} \ b_{21})^T \in \mathbb{R}^3. \quad (4.67)$$

Further, we identify anti-symmetric matrices \hat{b} as elements of the Lie algebra $\mathfrak{so}(3)$ of orthogonal group in 3-dimensions in the fundamental representation.

The fundamental matrix representation of an element $Y \in \mathfrak{se}(3)$ of the special Euclidean

transformations in 3-dimensions as

$$Y = \begin{pmatrix} 0 & \vec{0}^T \\ \vec{a} & \vec{m} \end{pmatrix} \quad (4.68)$$

where $\vec{a}, \vec{m} \in \mathbb{R}^3$ and $\vec{0} \in \mathbb{R}^3$, and we write this in short-hand notation as $Y = \{\vec{a}; \vec{m}\}$. Similarly, for $Z \in \mathfrak{se}(3)^*$, we write

$$Z := \{\vec{b}; \vec{n}\}^* = \begin{pmatrix} 0 & \vec{b}^T \\ \vec{0} & \vec{n}^T \end{pmatrix}. \quad (4.69)$$

Throughout the text we will often differentiate scalars with respect to vectors and matrices. We carry out matrix derivatives using the numerator-layout convention. For a matrix $X \in \mathbb{R}^{p \times q}$ and $y : \mathbb{R}^{p \times q} \rightarrow \mathbb{R}$ a scalar function $y(X)$, then

$$\frac{\partial y}{\partial X} = \begin{pmatrix} \frac{\partial y}{\partial X_{11}} & \frac{\partial y}{\partial X_{21}} & \cdots & \frac{\partial y}{\partial X_{p1}} \\ \frac{\partial y}{\partial X_{12}} & \frac{\partial y}{\partial X_{22}} & \cdots & \frac{\partial y}{\partial X_{p2}} \\ \vdots & \ddots & \vdots & \\ \frac{\partial y}{\partial X_{1q}} & \frac{\partial y}{\partial X_{2q}} & \cdots & \frac{\partial y}{\partial X_{pq}} \end{pmatrix} \quad (4.70)$$

. For $\vec{x} \in \mathbb{R}^d$ and $y : \mathbb{R}^{p \times q} \rightarrow \mathbb{R}$ a scalar function $y(\vec{x})$, then

$$\frac{\partial y}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_p} \end{pmatrix}. \quad (4.71)$$

Chapter 5

Cosserat rod models and geometric mechanics

We begin our study of geometric mechanics with the Cosserat rod. At any material point along its length, we will identify the configuration space of the material point with group of translations and rotations on three-dimensional Euclidean space $SE(3)$. We then use the Lie group-Lie algebra correspondence to construct a Maurer-Cartan form ξ that encapsulates the kinematic configuration of the Cosserat rod. By imposing integrability conditions on ξ , we find kinematic equations of motion defined on the Lie algebra. We then proceed to derive the dynamical equations of motion are then found using the *Euler-Poincaré theorem*. General non-conservative and non-constitutive dynamics are then derived from a generalised Lagrange-D'Alembert principle. As a result, we formulate the kinematics and general dynamics of a Cosserat rod, expressed in terms of its intrinsic geometry. Borrowing a term from robotics, we call the combined study of kinematics and dynamics *kinodynamics* [157, 158]. The formulation of conservative and constitutive kinodynamics of a Cosserat rod have been presented previously in [210, 211]. In Sec. 5.3 we then illustrate the usage of the Cosserat rod in physics in some example settings. In Sec. 5.4 we study *filaments*, which are formulated as Cosserat rods under kinematic constraints. The results presented here on the filament, as well as the Cosserat rod, will serve as paradigmatic examples of the generalised geometrised mechanics which is the main topic of discussion in Ch. 6.

Before we study the kinodynamics of Cosserat rods, we first consider the case of a free rigid body which undergo translations and rotations. We will first formulate the Lagrangian mechanics of the rigid body, expressing it in terms of its kinematic configuration space $SE(3)$, the group of special Euclidean transformations. We then construct a reduced Lagrangian density that take values in the corresponding Lie algebra $se(3)$. The canonical example of this procedure is the rotating free rigid body and was introduced in [42, 212, 213], which we here generalise here to include translation. We can consider the rigid body,

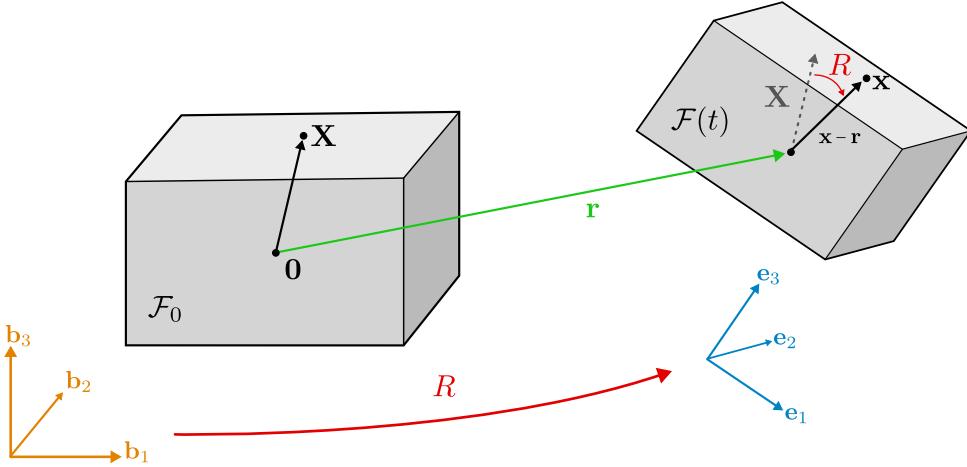


Figure 5.1: Illustration of the mapping from the reference configuration \mathcal{F}_0 to the current configuration $\mathcal{F}(t)$ of a rigid body. The configuration of the rigid body is specified via a translation of the center-of-mass \mathbf{r} (green), and rotation R (red), relative to a reference configuration with a center-of-mass located at $\mathbf{0}$. Material points \mathbf{X} in the reference configuration are thus mapped to $\mathbf{x}(\mathbf{X}) = \mathbf{r} + R\mathbf{X}$. An observer that sits on the rigid body and moves with it will have a frame $E = (\mathbf{e}_1 \mathbf{e}_2 \mathbf{e}_3)$ (blue), which rotates relative to the fixed frame of a still observer $B = (\mathbf{b}_1 \mathbf{b}_2 \mathbf{b}_3)$ (orange) as $E = RB$.

which is a finite-dimensional system, a ‘Cosserat point’. Therefore, it will serve as a useful foundation for the study of Cosserat rods.

5.1 Geometric mechanics of a rigid body

We consider mechanics in \mathbb{R}^3 , starting from the point-of-view of some observer with a *fixed* or *spatial* frame $B = (\mathbf{b}_1 \mathbf{b}_2 \mathbf{b}_3) \in \mathbb{R}^{3 \times 3}$. As we are working in Euclidean space, we will make use of the isomorphism $\mathbb{R}^3 \cong T\mathbb{R}^3$. Vectors $\mathbf{v} \in \mathbb{R}^3$ will be expressed in terms of their components in the fixed frame as $\mathbf{v} = \tilde{v}_i \mathbf{b}_i$, where the tilde denotes the fixed frame.

Let $\mathcal{M} \subset \mathbb{R}^3$ be the *reference configuration* of a rigid body, let $\mathbf{X} \in \mathcal{M}$ denote the *material coordinates* of \mathcal{M} , and let $\rho_0^V(\mathbf{X})$ be the mass density per unit material volume of the rigid body at \mathbf{X} . We define the material coordinates to be in the centre-of-mass frame, such that $\int_{\mathcal{M}} \rho_0^V(\mathbf{X}) X_i d^3 X = 0$, $i = 1, 2, 3$.

At time t the location of the material point at \mathbf{X} is given by $\mathbf{x}(\mathbf{X}, t)$. The latter can be related to the former as

$$\mathbf{x}(\mathbf{X}, t) = \mathbf{r}(t) + R(t)\mathbf{X}, \quad \mathbf{X} \in \mathcal{M} \quad (5.1)$$

where $\mathbf{r}(t)$ is the location of the centre-of-mass and $R(t)$ is the rotation of the rigid body relative to its reference configuration. Now consider the second term in Eq. 5.1,

$$R\mathbf{X} = \tilde{X}_i \mathbf{e}_i \quad (5.2)$$

where $\mathbf{e}_i = R\mathbf{b}_i$ is the *moving* or *body* frame, which we write as $E = (\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3) \in \mathbb{R}^{3 \times 3}$ and satisfies $E = RB$, and is illustrated in Fig. 5.1. The configuration of the rigid body can thus be written as a matrix

$$\mathcal{F}(t) = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{r}(t) & E(t) \end{pmatrix} \quad (5.3)$$

and we can relate it to the reference configuration as

$$\mathcal{F}(t) = \mathcal{F}_0 \Phi(t) \quad (5.4)$$

where

$$\mathcal{F}_0 = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & B \end{pmatrix} \quad (5.5)$$

is the reference configuration and

$$\Phi(t) = \begin{pmatrix} 1 & \mathbf{0}^T \\ B^T \mathbf{r} & B^T R(t) B \end{pmatrix} \quad (5.6)$$

is for each $t \in [0, T]$ an element of the special Euclidean group $SE(3)$ of translations and rotations. We will thus henceforth identify the configuration of the rigid body with Φ , up to a global rotation of \mathcal{F}_0 .

The form of Eq. 5.6 merits some explanation. $B^T \mathbf{r}$ can be seen as the component of coefficients of \mathbf{r} expressed in the fixed frame B , and $B^T R B$ is then the rotation R in that same basis. We therefore introduce the notation $\tilde{\mathbf{v}} = B^T \mathbf{v}$ for vectors $\mathbf{v} \in \mathbb{R}^3$. Similarly we write

$$\vec{v} = E^T \mathbf{v} \quad (5.7)$$

where \vec{v} is a vector $\mathbf{v} \in \mathbb{R}^3$ expressed in the moving frame E , with components $v_i = \mathbf{e}_i \cdot \mathbf{v}$. We thus have that $\mathbf{v} = v_i \mathbf{e}_i = \tilde{v}_i \mathbf{d}_i$.

We now derive the kinematic equations of motion of the rigid body, parametrised as a point $\Phi(t)$ in the Lie group $SE(3)$. The ‘velocity’ of the rigid body is $\dot{\Phi}(t) \in T_{\Phi(t)}SE(3)$, a vector in the tangent space at $\Phi(t)$, which incorporates both the translation of the centre-of-mass and the rotation of the frame. As $SE(3)$ is a Lie group, there is a natural map from $T_{\Phi(t)}SE(3)$ to its Lie algebra $\mathfrak{se}(3) \cong T_e SE(3)$ by left (or right)-translation

$$\Phi^{-1} \dot{\Phi} = N \in \mathfrak{se}(3) \quad (5.8)$$

where $\dot{\Phi} := \partial_t \Phi$. Equation 5.8 allows us to express the kinematics of the rigid body in terms of generalised velocities all taking values in the same Lie algebra. Equation 5.8 is an ordinary matrix differential equation and can be solved to find $\Phi(t)$ given a velocity $Y(t)$.

We now derive the kinematic equations of motion of the rigid body in terms of more familiar quantities. Let \vec{V} be the translational velocity of the center-of-mass, and $\vec{\Omega}$ the angular velocity of the rigid body, expressed in the moving frame basis. We write N as

$$N = \begin{pmatrix} 0 & \vec{0}^T \\ \vec{V} & \hat{\vec{\Omega}} \end{pmatrix} \quad (5.9)$$

where $\hat{\vec{\Omega}} \in \mathfrak{so}(3)$ is the angular velocity under the hat-map Eq. 4.66, which is the isomorphism between 3-vectors and 3×3 antisymmetric matrices, and where $\mathfrak{so}(3)$ is the Lie algebra of the special orthogonal group $SO(3)$. We can see N as a generalised velocity, defined on the Lie algebra. The inverse of Eq. 5.6 is given by

$$\Phi^{-1} = \begin{pmatrix} 1 & \mathbf{0}^T \\ -B^T R^T \mathbf{r} & B^T R^T B \end{pmatrix}. \quad (5.10)$$

We now evaluate the left-hand side of Eq. 5.8 to get

$$\begin{aligned} \Phi^{-1} \dot{\Phi} &= \mathcal{F}^{-1} \dot{\mathcal{F}} = \begin{pmatrix} 0 & \vec{0}^T \\ B^T R^T \dot{\mathbf{r}} & B^T R^T \dot{R} B \end{pmatrix} \\ &= \begin{pmatrix} 0 & \vec{0}^T \\ E^T \dot{\mathbf{r}} & E^T \dot{E} \end{pmatrix} \end{aligned} \quad (5.11)$$

Equating Eq. 5.9 and Eq. 5.11, we get the kinematic equations of motion

$$\dot{\mathbf{r}} = \mathbf{V} \quad (5.12a)$$

$$\dot{E} = E \hat{\vec{\Omega}}. \quad (5.12b)$$

which describes the translational motion of the centre-of-mass and the rotation of the frame E , respectively. Equation 5.12b can also be written as

$$\begin{aligned} \dot{\mathbf{e}}_j &= \mathbf{e}_i \hat{\Omega}_{ij} \\ &= \boldsymbol{\Omega} \times \mathbf{e}_i. \end{aligned} \quad (5.13)$$

where $\boldsymbol{\Omega} = \boldsymbol{\Omega}_i \mathbf{e}_i$. Eq. 5.13 provides a clear interpretation of the angular velocity. At time t the i th component of the angular velocity in the moving frame $\boldsymbol{\Omega}_i(t)$ gives the rate at which the rigid body rotates around the axis \mathbf{e}_i .

We note that the derivations above would simplify to some extent if we let the fixed frame equal the identity matrix $B = \mathbb{1}$, in which case $E = R$. In this section we will keep the distinction so as to make the distinction clear between the moving frame E and the fixed frame B . It is clear that \mathcal{F} (and \mathcal{F}_0) could also be considered Lie group-valued, and

indeed $\Phi^{-1}\dot{\Phi} = \mathcal{F}^{-1}\dot{\mathcal{F}}$. In the subsequent section we will make this identification.

Now we will consider the dynamics of the rigid body. We will start with the Euler-Lagrange equations defined on the Lie group, and then show the corresponding action principle defined on the Lie algebra. The kinetic energy of the rigid body is given by

$$\mathcal{K}(\dot{\Phi}) = \frac{1}{2} \int_{\mathcal{M}} \rho_0^V(\mathbf{X}) |\dot{\mathbf{x}}(\mathbf{X})|^2 d^3 X \quad (5.14)$$

which is here written explicitly as a function of $\dot{\Phi}$. To see this note that

$$\begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} = \mathcal{F}_0 \Phi \mathcal{F}_0^{-1} \begin{pmatrix} 1 \\ \mathbf{X} \end{pmatrix} \quad (5.15)$$

and so \mathbf{x} can be seen as a function of Φ . The dynamical equations of motions of a free rigid body with Lagrangian density $L : SE(3) \times TSE(3) \rightarrow \mathbb{R}$

$$\mathcal{L}(\Phi, \dot{\Phi}) = \mathcal{K}(\dot{\Phi}) \quad (5.16)$$

are found by invoking *Hamilton's principle*

$$\delta \int_0^T \mathcal{L}(\Phi(t), \dot{\Phi}(t)) dt = 0 \quad (5.17)$$

under variations $\Phi(t) \rightarrow \Phi(t) + \delta\Phi(t)$, where $\delta\Phi(t)$ is a variational test function which must vanish at the temporal boundaries, and where T is the upper-bound of the time-domain considered. The resulting equations of motion will be second-order equations for Φ in time. This formulation is often undesirable, as in numerical simulations errors accrue such as to push Φ out of the sub-manifold $SE(3) \subset \mathbb{R}^{4 \times 4}$. We will next formulate the corresponding equations of motion on the Lie algebra instead.

We have that

$$\mathcal{K}(\dot{\Phi}) = \frac{1}{2} m |\dot{\mathbf{r}}|^2 + \frac{1}{2} \int_{\mathcal{M}} \rho_0^V(\mathbf{X}) |\dot{R}\mathbf{X}|^2 d^3 X \quad (5.18)$$

where we have used that $\int_{\mathcal{M}} \rho_0^V(\mathbf{X}) \mathbf{X} d^3 X = 0$. Now, since

$$|\dot{R}\mathbf{X}| = |\dot{E}B^T X| = |\dot{E}\tilde{X}| = |E^T \dot{E}\tilde{X}| = |\hat{\Omega}\tilde{X}| = |\vec{\Omega} \times \tilde{X}| \quad (5.19)$$

we see that the rotational kinetic energy is a quadratic form in terms of the angular velocity $\vec{\Omega}$. Furthermore, as $|\dot{\mathbf{r}}| = |\vec{V}|$ we can write

$$\mathcal{K}(N) = \frac{1}{2} m |\vec{V}|^2 + \frac{1}{2} \vec{\Omega}^T \mathbb{I} \vec{\Omega} \quad (5.20)$$

where $\mathbb{I} \in \mathbb{R}^{3 \times 3}$ is the *moment of inertia* tensor, which is, unless the body is incidental to

a line, positive-definite. Eq. 5.20 is now a function of the Lie algebra-valued $N \in \mathfrak{se}(3)$. We define the *reduced Lagrangian density* $\ell : \mathfrak{se}(3) \rightarrow \mathbb{R}$

$$\ell(N) = \mathcal{K}(N) \quad (5.21)$$

which coincides with the regular Lagrangian density $\mathcal{L}(\Phi, \dot{\Phi}) = \ell(N)$ when $\Phi^{-1}\dot{\Phi} = N$. To find the corresponding Hamilton's principle

$$\delta \int_0^T \ell(N) dt = 0 \quad (5.22)$$

defined on the Lie algebra, we identify the space of admissible variations δN in terms of the variations on the Lie group-level $\delta\Phi$. We vary Eq. 5.8 to find

$$\delta N = -\Phi^{-1}\delta\Phi\Phi^{-1}\dot{\Phi} + \Phi^{-1}\delta\dot{\Phi}. \quad (5.23)$$

where we used $\delta(\Phi\Phi^{-1}) = 0$ to find $\delta\Phi^{-1}$. We define the variational test function $\eta = \Phi^{-1}\delta\Phi$, which satisfies $\dot{\eta} = -N\eta + \Phi^{-1}\delta\dot{\Phi}$. We get

$$\begin{aligned} \delta N &= \dot{\eta} + [N, \eta] \\ &= \dot{\eta} + \text{ad}_N\eta. \end{aligned} \quad (5.24)$$

As we will compute the corresponding variational principle for the Cosserat rod in the subsequent section, we will not go through the detailed derivation here. If the variation Eq. 5.22 is computed using Eq. 5.24, the resulting equations of motion are

$$\frac{d}{dt} \frac{\partial \ell}{\partial N} = \text{ad}_N^* \frac{\partial \ell}{\partial N} \quad (5.25)$$

where $\text{ad}_N^* : \mathfrak{se}(3)^* \rightarrow \mathfrak{se}(3)^*$ is the dual of the adjoint action, defined on the dual $\mathfrak{se}(3)^*$ to the Lie algebra, and $\frac{\partial \ell}{\partial N}$ is the matrix partial derivative which is carried out as in Eq. 4.70. Equation 5.25 is the result of the *Euler-Poincaré theorem* [42, 202, 214], and is also thus referred to as the *Euler-Poincaré equation*. Further details, and its derivation in the continuum case, will follow in the subsequent sections. Evaluating Eq. 5.25 leads to

$$\dot{\vec{V}} = \vec{V} \times \vec{\Omega} \quad (5.26a)$$

$$\dot{\vec{\Omega}} = \mathbb{I}\vec{\Omega} \times \vec{\Omega}. \quad (5.26b)$$

Eq. 5.8 and Eq. 5.26 together fully specify the kinodynamics of the free rigid body respectively. Note that Eq. 5.26 is expressed in the moving frame, and reduces to $\dot{\vec{V}} = 0$ and $\dot{\vec{\Omega}} = 0$ in a non-moving frame.

We will now briefly consider the case of a rigid body under the influence of an external

force. As the more general case of a continuum of rigid bodies will be treated in the following section, we will leave out many mathematical details, and only state some results here to preface the results that will follow.

External forcing of the rigid body can be included by considering the more general *Lagrange-D'Alembert principle* [42], which is, in integral form

$$\delta \int_0^T \mathcal{L}(\Phi, \dot{\Phi}) dt + \int_0^T (\mathbf{T}(\Phi, \dot{\Phi}))(\delta\Phi) dt = 0 \quad (5.27)$$

where $\mathcal{L}(\Phi, \dot{\Phi})$ is the Lagrangian density of the free rigid body, and $\mathbf{T} \in T^*SE(3)$ is a covector, where its argument is the variational test function $\delta\Phi$. \mathbf{T} is here a generalised external force on the rigid body, containing both the force and the moment applied on the rigid body. As shown in [215], Eq. 5.27 in reduced form leads to

$$\delta \int_0^T \ell(N) dt + \int_0^T \langle T, \eta \rangle dt = 0 \quad (5.28)$$

where T is the generalised external force pulled-back onto the dual Lie algebra $\mathfrak{se}(3)^*$, η the variational test function defined on the Lie algebra, and $\langle \cdot, \cdot \rangle : \mathfrak{se}(3)^* \times \mathfrak{se}(3) \rightarrow \mathbb{R}$ is an inner product. If we let the \vec{f} and \vec{m} be the external force and moment respectively, the resulting equations of motion are

$$\dot{\vec{V}} = \vec{V} \times \vec{\Omega} + \vec{f} \quad (5.29a)$$

$$\dot{\vec{\Omega}} = \mathbb{I}\vec{\Omega} \times \vec{\Omega} + \vec{m}. \quad (5.29b)$$

In the following section we will generalise these equations of motion for the Cosserat rod.

5.2 Geometric Cosserat rod mechanics

5.2.1 Cosserat rod kinematics

General Cosserat media can be described as a microstructured sub-manifold of Euclidean space. At each point-continua on the sub-manifold n directors are attached, where each director can in general vary in both magnitude and direction independently. The *special Cosserat rod* is an example of a Cosserat system where the one-dimensional continuum body is called the *center-line* $\mathbf{r}(u)$, with two orthogonal directors at each material point $u \in [0, L_0]$, where L_0 is the rest length of the rod, that are constrained to be *inextensible* (constant magnitude) and orthogonal. Equivalently, we can thus see the special Cosserat rod as a connected string of rigid body cross-sections. These cross-sections are represented as orthonormal frames we call the *material frames* $E(u) = (\mathbf{e}_1(u) \ \mathbf{e}_2(u) \ \mathbf{e}_3(u))$, which

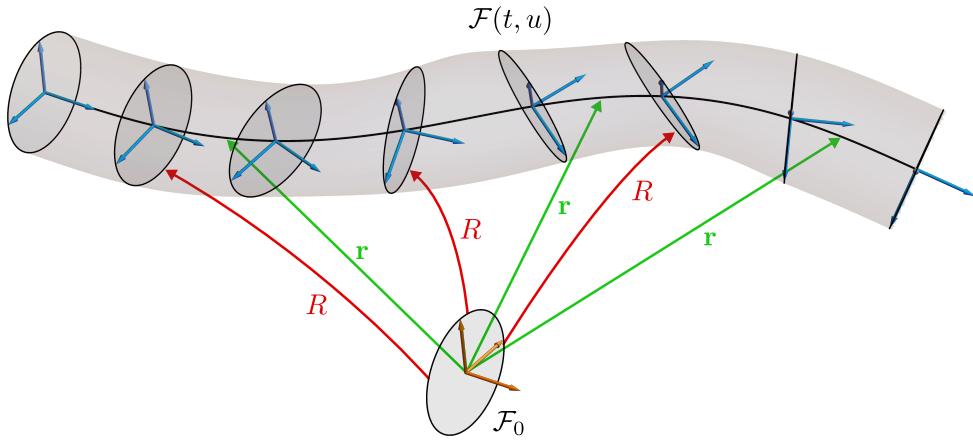


Figure 5.2: Illustration of the mapping from the reference configuration \mathcal{F}_0 to the current configuration $\mathcal{F}(t, u)$ of a Cosserat rod. The configuration of a Cosserat rod is specified by the center-line $\mathbf{r}(u)$, $u \in [0, L_0]$ and a rigid body cross section attached at each u , related to a reference configuration \mathcal{F}_0 , via translation (red) and rotation (red) respectively.

rotate relative to each other along the material coordinate u . A pair $\mathbf{r}(u), E(u)$) is known as a *trihedron*. See Fig. 4.1 for an illustration.

We will show that the kinematic configuration space of the Cosserat rod is the Lie group of special Euclidean transformations $SE(3)$, and then proceed to formulate its kinematic equations of motion on the corresponding Lie algebra $\mathfrak{se}(3)$.

Lie group–Lie algebra correspondence

The center-of-mass of the cross-section at u is given by $\mathbf{r}(u)$, and its orientation is specified by the material frame $E(u) = (\mathbf{e}_1(u) \ \mathbf{e}_2(u) \ \mathbf{e}_3(u))$, which is an orthonormal triad. As E is a basis for \mathbb{E}^3 , the vector of components in the material frame basis $\vec{v} = E^T \mathbf{v}$ will be said to be expressed in the *moving frame*. The configuration of the Cosserat rod is thus specified by the center-line and material frame, which we write as

$$\mathcal{F}(t, u) = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{r}(t, u) & E(t, u). \end{pmatrix} \quad (5.30)$$

We can relate rod configurations to a reference coordinate system as

$$\mathcal{F}(t, u) = \mathcal{F}_0 \Phi(t, u) \quad (5.31)$$

where

$$\mathcal{F}_0 = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & B \end{pmatrix} \quad (5.32)$$

is the reference coordinate system, and $\Phi(t, u) \in SE(3)$ is the transformation between the two. See Fig. 5.2 for an illustration. In contrast the previous section, here we set the fixed

frame to $B = \mathbb{1}_{3 \times 3}$, to simplify the notation. Such that $\mathcal{F}_0 = \mathbb{1}_{4 \times 4}$ and

$$\Phi(t, u) = (\mathbf{r}; E) = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{r}(t, u) & E(t, u) \end{pmatrix} \in SE(3). \quad (5.33)$$

where we have introduced the short-hand notation for $\Phi = (\mathbf{r}; E)$ for $SE(3)$ elements. We thus have that $\mathcal{F}(t, u) = \Phi(t, u)$, and we will henceforth identify the Cosserat rod configuration with the group element $\Phi(t, u)$. The kinematic motion of the Cosserat rod can thus be fully specified by a space-time sheet $\Phi(t, u) \in SE(3)$ in the Lie group of special Euclidean transformations. In other words, the spatio-temporal configuration of the Cosserat rod can be described as a 2-dimensional immersion $N \subset SE(3)$, defined by the map $\Phi : W \rightarrow SE(3)$, where we call $W = [0, T] \times [0, L_0]$ the *kinematic base space*, and $\mathcal{N} = \Phi(W)$. Therefore we call $SE(3)$ the *kinematic configuration space* of the Cosserat rod, and Φ the *spatio-temporal configuration* of the system. In other words, \mathcal{N} can be seen as a ‘sheet’ in $SE(3)$, and is parametrised by the function $\Phi : W \rightarrow \Phi(W) = \mathcal{N}$. The manifolds $[0, L_0]$ and $[0, T]$ will be called the *material base space* and the *time domain* respectively. Furthermore, we may draw a distinction between the *internal configuration space* $SO(3)$, corresponding to the material frame, and the *external configuration space* $\mathbb{E}^3 \cong SE(3)/SO(3)$, corresponding to the center-line.

At a point $\Phi(t, u) \in \mathcal{N}$, we can consider velocities in the temporal and spatial directions $\dot{\Phi}(t, u)$ and $\Phi'(t, u)$ respectively. Thus by differentiating the map Φ we get a vector field

$$d\Phi = \dot{\Phi}dt + \Phi'du \quad (5.34)$$

as discussed in Sec. 4.2.5, where $d\Phi(t, u) \in T_{\Phi(t, u)}\mathcal{N}$ for all $(t, u) \in W$. Here it is worth clarifying a technical point to avoid confusion. Although we say $d\Phi$ is a vector field on N , we have written it as a 1-form on the kinematic base space W . These two notions are compatible. Precisely, $d\Phi$ is a *$T\mathcal{N}$ -valued 1-form on W* . This is intuitive, as we would expect velocities and spatial rates-of-change on a continuum system to be ‘densities’ with respect to t and u respectively. Therefore, vector fields on the Cosserat rod are vector-valued differential forms.

Via left-translation, we can relate $d\Phi$ to a $\mathfrak{se}(3)$ -valued 1-form on W

$$\xi = \Phi^{-1}d\Phi \quad (5.35)$$

where $\xi = \Phi^*\omega$, the pull-back $\Phi^*\omega$ of the Maurer-Cartan ω form, as discussed in Sec. 4.2.6. ξ contains all geometric information contained in Φ , up to rigid body transformations, and as a Lie algebra-valued object ξ will turn out to be easier to work with than Φ or $d\Phi$. In what follows, we will formulate the kinematics of the Cosserat rod entirely within the Lie algebra using ξ .

Kinematic equations of motion

Equation 5.35 shows how to construct ξ from the Lie group function Φ . For our purposes it will be more relevant to go the other direction: Given a ξ , is there a corresponding Φ ? The answer can be formulated in terms of integrability. As a motivating and intuitive example from multi-variate calculus, consider a vector-valued function $\mathbf{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$. There exists a potential $U : \mathbb{R}^3 \rightarrow \mathbb{R}$ such that $\mathbf{F} = -\nabla \cdot U$ if and only if the differential $\phi = \frac{\partial F_i}{\partial x^i} dx^i$ is *exact*. In \mathbb{R}^3 , and for Lie groups, this is equivalent to $d\phi = 0$ ¹. In Sec. 4.2.6, we showed that there is a Φ for which the right-hand side of Eq. 5.35 holds true if and only if

$$d\xi + \xi \wedge \xi = 0. \quad (5.36)$$

This is an integrability condition on ξ , and is a Maurer-Cartan equation on the sub-manifold N . As we will show, Eq. 5.36 is in fact a concise expression of the kinematics of a Cosserat rod. We expand ξ in terms of its temporal and spatial components

$$\xi = Ndt + Xdu \quad (5.37)$$

where $X(t, u), N(t, u) \in \mathfrak{se}(3)$, which we write as

$$X = \{\vec{\theta}; \vec{\pi}\} = \begin{pmatrix} 0 & \vec{0}^T \\ \vec{\theta} & \hat{\pi} \end{pmatrix} \quad (5.38a)$$

$$N = \{\vec{V}; \vec{\Omega}\} = \begin{pmatrix} 0 & \vec{0}^T \\ \vec{V} & \hat{\Omega} \end{pmatrix} \quad (5.38b)$$

where $\vec{\theta}(t, u), \vec{\pi}(t, u), \vec{V}(t, u), \vec{\Omega}(t, u) \in \mathbb{R}^3$ are vectors expressed in the material frame basis, and where we have introduced the notation $X = \{\vec{\theta}; \vec{\pi}\}$ as a shorthand for matrices like Eq. 5.38a. We will call X the *spatial reconstruction field* and N the *generalised velocity field*. Following a similar derivation as in Eq. 5.11, we find

$$\Phi^{-1} d\Phi = \begin{pmatrix} 0 & \vec{0}^T \\ E^T d\mathbf{r} & E^T dE \end{pmatrix}. \quad (5.39)$$

From Eq. 5.38 we then find that

$$d\mathbf{r} = \mathbf{e}_i V_i dt + \mathbf{e}_i \theta_i du \quad (5.40a)$$

$$d\mathbf{e}_j = \mathbf{e}_i \hat{\Omega}_{ij} dt + \mathbf{e}_i \hat{\pi}_{ij} du \quad (5.40b)$$

¹This is known as the *Poincaré lemma* and holds true locally for arbitrary manifolds.

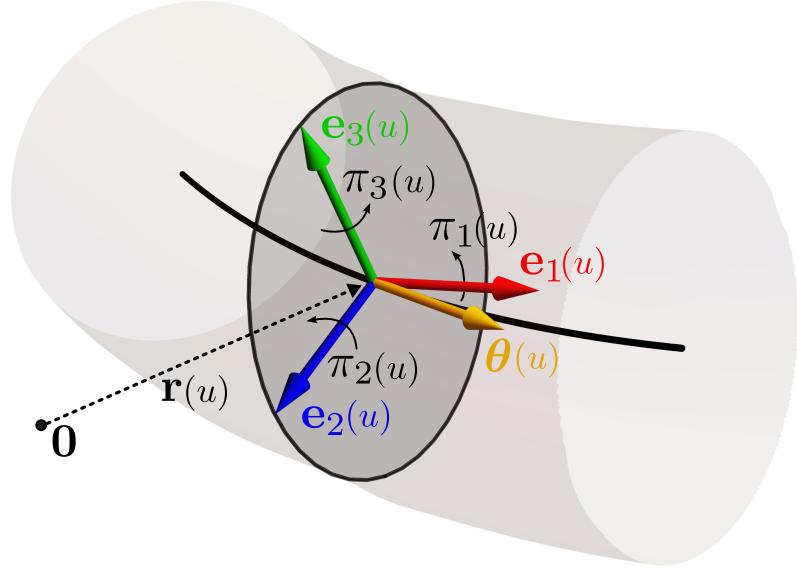


Figure 5.3: Depicts the rate-of-change along u of the material frame located at $\mathbf{r}(u)$ (dashed line). The spatial derivative of the center-line (solid black line) at u is $\boldsymbol{\theta}(u) = \mathbf{r}'(u)$ (yellow arrow), and π_1 , π_2 and π_3 is the angular rate-of-change of the material frame around \mathbf{e}_1 (red), \mathbf{e}_2 (blue) and \mathbf{e}_3 (green) respectively. The analogous description holds true for \vec{V} and $\vec{\Omega}$, but in time t rather than material coordinate u .

which we can also write as

$$d\mathbf{r} = \mathbf{V}dt + \boldsymbol{\theta}du \quad (5.41a)$$

$$dE = E\hat{\vec{\Omega}}dt + E\hat{\vec{\pi}}du. \quad (5.41b)$$

We can thus identify $\mathbf{V} = E\vec{V} = \partial_t \mathbf{r}$ as the translational velocity of the center-line, $\boldsymbol{\theta} = E\vec{\theta} = \partial_u \mathbf{r}$ as rate-of-change of the center-line along the material coordinate, and from Eq. 5.40b it follows that

$$\dot{\mathbf{e}}_i = \boldsymbol{\Omega} \times \mathbf{e}_i \quad (5.42a)$$

$$\mathbf{e}'_i = \boldsymbol{\pi} \times \mathbf{e}_i \quad (5.42b)$$

where $\dot{\mathbf{e}}_i := \partial_t \mathbf{e}_i$ and $\mathbf{e}'_i = \partial_u \mathbf{e}_i$, which show that $\boldsymbol{\Omega} = E\vec{\Omega}$ and $\boldsymbol{\pi} = E\vec{\pi}$ are the temporal and ‘spatial’ angular velocities of the material frame respectively. The spatial part of Eq.5.40 is illustrated in Fig. 5.3. We can thus see N as a generalised velocity in the Lie algebra. Equation 5.40b also shows how derivatives of vector functions $\mathbf{v}(t, u)$ can be

expressed in terms of $\vec{v}(t, u)$ in the moving frame basis. We have

$$\begin{aligned}
d\mathbf{v} &= d(\mathbf{e}_i v_i) \\
&= \mathbf{e}_i(dv_i) + (d\mathbf{e}_i)v_i \\
&= E(d\vec{v}) + (E\hat{\pi} + E\hat{\Omega})v_i \\
&= E(D_t \vec{v} dt + D_u \vec{v} du)
\end{aligned} \tag{5.43}$$

such that $E^T d\mathbf{v} = D_u \vec{v} du + D_t \vec{v} dt$, where

$$D_t = \partial_t + \hat{\Omega} \tag{5.44a}$$

$$D_u = \partial_u + \hat{\pi} \tag{5.44b}$$

are the spatial and temporal *material derivatives* respectively. We thus have that $E^T \mathbf{v}' = D_u \vec{v}$ and $E^T \dot{\mathbf{v}} = D_t \vec{v}$.

For a given spatio-temporal velocity field $N(t, u)$, Eq. 5.36 imposes a condition on the spatial reconstruction field $X(t, u)$. The resulting equations are the kinematic equations of motion. Substituting Eq. 5.37 into the left-hand side of Eq. 5.36

$$\begin{aligned}
d\xi + \xi \wedge \xi &= N' dt \wedge du + \dot{X} dt \wedge du + X N dt \wedge du + N X dt \wedge du \\
&= (\dot{X} - N' + [N, X]) dt \wedge du,
\end{aligned} \tag{5.45}$$

we get

$$\dot{X} = \mathcal{D}_u N. \tag{5.46}$$

where we have defined

$$\mathcal{D}_u = \partial_u + \text{ad}_X \tag{5.47}$$

which we will call a *generalised material derivative*. Equation 5.46 is the kinematic equation of motion for the Cosserat rod, which expresses the evolution of the spatial configuration X , in terms of the generalised velocity N . To express the kinematics in terms of the non-generalised velocities, we substitute Eq. 5.38 into Eq. 5.46 to find

$$D_t \vec{\theta} = D_u \vec{V} \tag{5.48a}$$

$$\partial_t \vec{\pi} = D_u \vec{\Omega}, \tag{5.48b}$$

which are equivalent to Eq. 5.46 and are the kinematic equations of motion expressed in the moving frame. At the Lie group-level the corresponding equations are

$$\dot{\mathbf{r}} = \mathbf{V} \tag{5.49a}$$

$$\dot{\mathbf{e}}_i = \boldsymbol{\Omega} \times \mathbf{e}_i \tag{5.49b}$$

From the numerical standpoint, the benefit of Eq. 5.48b over Eq. 5.49b is clear. Errors in integrating the latter can in general lead to $|\mathbf{e}_i| \neq 1$ and $\mathbf{e}_i \cdot \mathbf{e}_j \neq 0$, whilst by errors in Eq. 5.48b are by construction Lie algebra-valued. In other words, errors in Eq. 5.49b accrue in $\mathbb{R}^{3 \times 3}$, whilst errors in Eq. 5.48b accrue within the Lie group itself. Mathematically, this is due to the fact that Eq. 5.46 is defined on the linear space $W \times \mathfrak{se}(3)$, as opposed to Eq. 5.49 which is defined on the non-linear space $G \times TG$.

Types of deformations

Fig. 4.2 shows the various types of deformations that are kinematically possible for the Cosserat rod. Here we will describe how these deformations are encoded in $\vec{\theta}$ and $\vec{\pi}$. Since $\boldsymbol{\theta} = \mathbf{r}'$ we have that if $\vec{\theta}(u) \propto (1 \ 0 \ 0)$ then the material frame at u , pointing in the direction $\mathbf{e}_1(u)$ is aligned with the center-line. We thus see that $\vec{\theta}$ measures the *shear* of the rod. Note that this is because $\vec{\theta}$ is expressed in the moving frame; $\boldsymbol{\theta}$ does not contain any information about the shear.

Twist is the rotation of the material frame around \mathbf{e}_1 , and the rate of twist along the material coordinate is given by π_1 . Similarly, the rotation of the material frame around \mathbf{e}_2 and \mathbf{e}_3 are given by π_2 and π_3 respectively, as depicted in Fig. 5.3.

In a non-moving frame the *bending* of the center-line is encoded in \mathbf{r} , likewise $\boldsymbol{\theta}$. In the moving frame it is encoded in both $\vec{\theta}$ and $\vec{\pi}$ simultaneously. For instance, if $\vec{\theta}$ is constant along u and π_2 or π_3 are non-zero along u , then the rod is bending.

In general velocities \vec{V} will cause the Cosserat to *extend*. Locally, the extension is described by the scalar $h(u) = |\mathbf{r}'| = |\vec{\theta}|$, which can be seen as the square root of the metric on the center-line induced by the Euclidean metric on \mathbb{R}^3 . Using h we can define an arc-length coordinate as

$$ds = h(u)du \quad (5.50)$$

which relates the arc-length increment ds to the material length increment du . In the arc-length parametrisation we thus have $|\frac{\partial \mathbf{r}}{\partial s}| = 1$ for all $s \in [0, L]$ where

$$L[h(u)] = \int_0^{L_0} h(u)du \quad (5.51)$$

is the total arc-length of the center-line. $h(u)$ thus measures the local extension of the rod, where $h(u) = |\vec{\theta}| = 1$ indicates that the rod is not suffering an extension at u .

We define a unit-vector tangent to the center-line $\mathbf{t} = \boldsymbol{\theta}/|\boldsymbol{\theta}|$. The *curvature* of the center-line is then a scalar defined as

$$\kappa = |\mathbf{t}'|. \quad (5.52)$$

In the literature it is more common to define the curvature in terms of the arc-length

parametrisation, which we write as $\tilde{\kappa} = |\partial_s \mathbf{t}| = |\partial_s^2 \mathbf{r}|$.

Reconstructing the Cosserat rod from ξ

Solving Eq. 5.46, for a given velocity N , yields the Lie algebra-valued 1-form ξ which contains, in infinitesimal form, information from which the Cosserat rod \mathcal{F} can be reconstructed. Firstly, recall that we can reconstruct Φ , the representation of the Cosserat rod as a space-time sheet in the Lie group, from ξ using Eq. 5.35. Secondly, from Eq. 5.31 we note that $\Phi^{-1}d\Phi = \mathcal{F}^{-1}d\mathcal{F}$. Therefore we can reconstruct the Cosserat rod from ξ by solving the equation

$$d\mathcal{F} = \mathcal{F}\xi. \quad (5.53)$$

We will now put the above into a form solvable as a matrix ordinary differential equation. By expanding $\mathcal{F}^{-1}d\mathcal{F}$ into its components, and from Eq. 5.37 we find that

$$\mathcal{F}^{-1}\mathcal{F}' = X \quad (5.54a)$$

$$\mathcal{F}^{-1}\dot{\mathcal{F}} = Y. \quad (5.54b)$$

For fixed $t = \bar{t}$ and given initial conditions Eq. 5.54a can be used to find $\mathcal{F}(u, \bar{t})$ for $u \in [0, L_0]$, and for fixed $u = \bar{u}$ and given initial conditions Eq. 5.54b can be used to find $\mathcal{F}(\bar{u}, t)$ for $t \in [0, T]$. We can therefore trace out and reconstruct $\mathcal{F}(t, u)$ by repeatedly solving Eqs. 5.54a and 5.54b. Due to Eq. 5.36, which ensures that $d\mathcal{F}$ is an ‘exact’ differential, the resulting $\mathcal{F}(t, u)$ is invariant with respect to the particular path in (t, u) -space taken to reconstruct it. Therefore, a suitable scheme to reconstruct \mathcal{F} is to first reconstruct the Cosserat rod in time along $u = 0$ by solving

$$\bar{\mathcal{F}}(t) = \bar{\mathcal{F}}(t)\dot{\bar{\mathcal{F}}}(t), \quad (5.55)$$

with initial condition $\bar{\mathcal{F}}(0) = \mathcal{F}^{0,0}$, where the latter corresponds to choosing values for $\mathbf{r}(0, 0)$ and $E(0, 0)$. Secondly, for any given time $t = \bar{t}$ we can reconstruct the Cosserat rod in space by solving

$$\mathcal{F}(u, \bar{t}) = \mathcal{F}(u, \bar{t})\mathcal{F}(u, \bar{t})', \quad (5.56)$$

with initial condition $\mathcal{F}(0, \bar{t}) = \bar{\mathcal{F}}(\bar{t})$. See Appendix C.1 for a detailed description of the numerical algorithm.

Closed Cosserat rods

Thus far we have considered *open* Cosserat rods, for which $\mathbf{r}(0) \neq \mathbf{r}(L_0)$ and $E(0) \neq E(L_0)$. For *closed* Cosserat rods the material coordinate $u \in [0, L_0]$ is periodic in its domain, and the center-line and material frame must also be periodic and continuous in u . Although

seemingly complicated, the following derivations will result in showing that the case of a closed Cosserat can be in practice treated equivalently to an open Cosserat rod.

A closed rod implies that

$$X(0) = X(L_0). \quad (5.57)$$

However, Eq. 5.57 is only a necessary, but not sufficient, condition for the rod to close. The sufficient condition is for the constraint

$$\mathcal{F}(0, t) = \mathcal{F}(L_0, t) \quad (5.58)$$

to hold for all t . This is an integral constraint on X , as formally we can write

$$\mathcal{F}(L_0, t) = \mathcal{F}(0, t) \mathcal{U} \exp \left\{ \int_0^{L_0} X du \right\} \quad (5.59)$$

where \mathcal{U} denotes a u -ordered integral. The integral constraint on X is therefore

$$\mathcal{U} \exp \left\{ \int_0^{L_0} X du \right\} = \mathbb{1}. \quad (5.60)$$

However, if N is a smooth function of u , the kinematic equations of motion will preserve Eq. 5.60. This means that if the rod is initialised to obey Eq. 5.60 at $t = 0$, it will continue to do so for all t .

In practice, in order to simulate a closed filament we must thus first construct on the Lie group-level a periodic $\mathcal{F}(u, 0)$, and then use Eq. 5.54a to find $X(u, 0)$ that obeys Eq. 5.60.

5.2.2 Cosserat rod dynamics

Constitutive dynamics

Thus far we have considered the Cosserat rod abstractly as a sub-manifold of $SE(3)$. Physically, we can see the the Cosserat rod as a kinematic approximation of a deformable and slender three-dimensional rod. Let $\mathcal{M} = D \times [0, L_0] \subset \mathbb{R}^3$ be the reference configuration of such a rod, where $D \subset \mathbb{R}^2$ is the cross-section of the slender rod and is of arbitrary shape. Let $\mathbf{X} \in \mathcal{M}$ denote the material coordinates of \mathcal{M} , such that $X_1 \in [0, L_0]$ and $(X_2, X_3) \in \mathcal{D}$ and for simplicity let ρ_0^V be a constant mass density per unit material volume at each \mathbf{X} . At time t the location of the material point at \mathbf{X} is given by $\mathbf{x}(\mathbf{X}, t)$, which are the deformed coordinates. We define (X_2, X_3) to be in the centre-of-mass of the cross-section, such that $\int_{\mathcal{D}} X_\gamma dX_2 dX_3 = 0$, $\gamma = 2, 3$.

The Cosserat rod models the slender rod under the kinematic assumption that the material fibres that run radially from the center of \mathcal{D} to its edge are fixed. In other words

we impose that

$$\mathbf{x}(\mathbf{X}, t) = \mathbf{r}(t, u) + X_\gamma \mathbf{e}_\gamma(t, u), \quad \gamma = 2, 3 \quad (5.61)$$

where $\mathbf{r}(t, u)$ is the center-line of the rod and where $u = X_1$ is the material coordinate along the center-line. The three-dimensional rod has thus been replaced with the Cosserat rod, which only has one spatial dimension. In this coarse-graining, the two missing spatial dimensions have been replaced by rigid body cross-sections attached at each $u \in [0, L_0]$, with orientation specified by the two directors \mathbf{e}_2 and \mathbf{e}_3 .

As in Sec. 5.1 we formulate the dynamics of the Cosserat rod via a Lagrangian formulation. We will again be using a reduction of the Lagrangian to construct dynamics defined on the Lie algebra. We derive the constitutive dynamics by first computing the kinetic energy of the rod in terms of its generalised velocity $N = \{\vec{V}; \vec{\Omega}\}$.

We assume that the dynamics is defined by a Lagrangian density of the form $\mathcal{L}(\Phi, d\Phi)$, and that its kinetic energy term is given by

$$\int_0^{L_0} \mathcal{K}(\dot{\Phi}) du = \frac{1}{2} \int_{\mathcal{M}} \rho_0^V |\dot{\mathbf{x}}|^2 d^3X \quad (5.62)$$

where \mathcal{K} is the kinetic energy density along the material coordinate u . Substituting in Eq. 5.61 we get

$$\int_0^{L_0} \mathcal{K}(\dot{\Phi}) du = \frac{1}{2} \int_0^{L_0} \rho_0 |\mathbf{V}|^2 du + \frac{1}{2} \rho_0^V \int_{\mathcal{M}} \mathbf{V} \cdot \dot{\mathbf{e}}_\gamma X_\gamma d^3X + \frac{1}{2} \rho_0^V \int_{\mathcal{M}} |X_\gamma \dot{\mathbf{e}}_\gamma|^2 d^3X \quad (5.63)$$

where $\rho_0 = A \rho_0^V$ is the mass density per unit material length, and $A = \int_D dX_2 dX_3$ is the area of the cross-section, and we have used $\dot{\mathbf{r}} = \mathbf{V}$. The second term in Eq. 5.63 vanishes when the integral over D is evaluated. To evaluate the third term, let $\bar{\mathbf{X}} = X_\gamma \mathbf{e}_\gamma$, and note that $\mathbf{e}_i = \boldsymbol{\Omega} \times \mathbf{e}_i$, and so

$$|X_\gamma \dot{\mathbf{e}}_\gamma|^2 = |X_\gamma \boldsymbol{\Omega} \times \mathbf{e}_\gamma|^2 = |\boldsymbol{\Omega} \times \bar{\mathbf{X}}|^2 = |E(\vec{\Omega} \times \vec{\bar{X}})|^2 = |\vec{\Omega} \times \vec{\bar{X}}|^2 \quad (5.64)$$

where we used $(E\vec{\Omega}) \times (E\vec{\bar{X}}) = E(\vec{\Omega} \times \vec{\bar{X}})$. As $|(\vec{\Omega} \times \vec{\bar{X}})|^2$ is a quadratic form in $\vec{\Omega}$, we can write

$$\frac{1}{2} \rho_0^V \int_{\mathcal{M}} |X_\gamma \dot{\mathbf{e}}_\gamma|^2 d^3X = \frac{1}{2} \int_0^{L_0} \vec{\Omega}^T \mathbb{I} \vec{\Omega} du \quad (5.65)$$

where \mathbb{I} is the cross-sectional moment-of-inertia per unit material length. We can thus write the kinetic energy of the Cosserat rod as $K(N) = \int_0^{L_0} \mathcal{K}(N) du$ where

$$\mathcal{K}(N) = \frac{1}{2} \rho_0 |\vec{V}|^2 + \frac{1}{2} \vec{\Omega}^T \mathbb{I} \vec{\Omega} \quad (5.66)$$

is the kinetic energy per material unit length, which is now explicitly a function of the generalised velocity N .

Having established that the kinetic energy of the Cosserat rod has a reduced form, we now assume that we have a reduced Lagrangian density of the form

$$\ell(\xi) = \mathcal{K}(N) - \mathcal{U}(X) \quad (5.67)$$

where $\mathcal{U}(X)$ is the potential energy per material unit length of a given spatial configuration $X = \{\vec{\theta}; \vec{\pi}\}$, and $\ell(\xi)$ is here a Lagrangian density defined on the Lie algebra. The reduced Lagrangian density satisfies $\ell(\xi) = \mathcal{L}(\Phi, d\Phi)$ when $\xi = \Phi^{-1}d\Phi$.

The reduced Lagrangian density Eq. 5.67 is explicitly invariant under Galilean transformations, as it is written in terms of the invariant quantities X and N . Physically, the fact that a Lagrangian $\mathcal{L}(\Phi, d\Phi)$ admits a reduction $\ell(\xi)$ entails that only differential deformations couple to the dynamics. In other words the Lagrangian is really only a function of the tangent space $\mathcal{L}(d\Phi)$. We will henceforth omit the argument in Φ when the Lagrangian density admits a reduced form.

The *generalised conjugate momentum* of the Cosserat rod is

$$S := \frac{\partial \ell}{\partial N} = \frac{\partial \mathcal{K}}{\partial N} = \begin{pmatrix} 0 & \vec{P}^T \\ \vec{0} & \hat{L}^T \end{pmatrix} \in \mathfrak{se}(3)^* \quad (5.68)$$

where $\vec{P} = \rho_0 \vec{V}$ and $\vec{L} = \mathbb{I} \vec{\Omega}$. The matrix derivative is carried out using the numerator-layout convention Eq. 4.70, such that $S_{ij} = \partial \ell / \partial N_{ji}$. The matrix representation of the dual Lie algebra can thus be seen as the space $\mathfrak{se}(3)^* = \{Y^T : Y \in \mathfrak{se}(3)\}$. It will become clear that \vec{P} is the linear momentum of the center-line per unit material length, and \vec{L} is the angular momentum per unit material length of the material frame. We use the short-hand $S = \{\vec{P}; \vec{L}\}^*$, corresponding to matrices of the form Eq. 5.68. The matrix derivative $\frac{\partial \ell}{\partial N}$ is evaluated only with respect to the non-zero components of N , for which we have $S_{ji} = \frac{\partial \ell}{\partial N_{ij}}$, and $S_{ji} = 0$ otherwise. S can be seen to take values in the dual Lie algebra $\mathfrak{se}(3)^*$, which is the space of linear operators $\mathfrak{se}(3) \rightarrow \mathbb{R}$ on the Lie algebra. To see this, note that $\mathcal{K}(N)$ can be written as

$$\mathcal{K}(N) = \frac{1}{2} \langle N, P \rangle \quad (5.69)$$

where $\langle \cdot, \cdot \rangle : \mathfrak{se}(3) \times \mathfrak{se}(3)^* \rightarrow \mathbb{R}$ is an inner product defined as

$$\langle A, B \rangle = \vec{a} \cdot \vec{b} + \vec{m} \cdot \vec{n} \quad (5.70)$$

where $A = \{\vec{a}; \vec{m}\} \in \mathfrak{se}(3)$ and $B = \{\vec{b}; \vec{n}\} \in \mathfrak{se}(3)^*$. Similarly, we can define a *generalised stress*, conjugate to the spatial configuration X , as

$$Q := \frac{\partial \mathcal{U}}{\partial X}. \quad (5.71)$$

We write its components as $Q = \{\vec{F}; \vec{M}\}^*$ where

$$F_i = \frac{\partial \mathcal{U}}{\partial \theta_i} \quad (5.72a)$$

$$M_i = \frac{\partial \mathcal{U}}{\partial \pi_i} \quad (5.72b)$$

are, as will be shown, the force on the center-line and the moment on the material frame respectively.

To find dynamic equations of motion for P and Q , we invoke Hamilton's principle that variations

$$\delta \int_W \ell(\xi) dt \wedge du \quad (5.73)$$

must vanish, where $\ell(\xi)$ is again the reduced Lagrangian density for the system. As for the rigid body in Sec. 5.1, we must first derive the permissible form of the variations. We first consider the Lagrangian density $\mathcal{L} : SE(3) \times TSE(3) \rightarrow \mathbb{R}$ defined on the Lie group, which satisfies $\mathcal{L}(\Phi, d\Phi) = \ell(\xi)$ when $\Phi^{-1}d\Phi = \xi$. The Euler-Lagrange equations are found by imposing

$$\delta \int_W \mathcal{L}(\Phi, d\Phi) dt \wedge du = 0 \quad (5.74)$$

under variations $\Phi \rightarrow \Phi + \delta\Phi$, where $\delta\Phi(u, t) \in TSE(3)$ is a variational test function which must vanish at the temporal boundaries. We will now assume that the Lagrangian density can be written as a function $\mathcal{L} = \mathcal{L}(d\Phi)$, purely in terms of the tangent vectors $d\Phi$. Repeating similar steps as in Sec. 5.1, we vary $\xi = \Phi^{-1}d\Phi$ to get

$$\delta\xi = -\Phi^{-1}\delta\Phi\Phi^{-1}d\Phi + \Phi^{-1}\delta(d\Phi). \quad (5.75)$$

We define the variational test function $\eta = \Phi^{-1}\delta\Phi$, which satisfies $d\eta = -N\eta + \Phi^{-1}\delta(d\Phi)$, to get

$$\begin{aligned} \delta\xi &= d\eta + \text{ad}_\xi\eta \\ &= (\dot{\eta} + \text{ad}_N)dt + (\eta' + \text{ad}_X\eta)du \\ &= \delta X du + \delta N dt \end{aligned} \quad (5.76)$$

We now proceed to evaluate Eq. 5.73. We first note that since

$$\begin{aligned} \delta\ell &= \frac{\partial\ell}{\partial X_{ij}}\delta X_{ij} + \frac{\partial\ell}{\partial N_{ij}}\delta N_{ij} \\ &= \langle S, \delta N \rangle - \langle Q, \delta X \rangle, \end{aligned} \quad (5.77)$$

we have

$$\begin{aligned}
\delta \int_W \ell(\xi) dt \wedge du &= \int_W \{\langle \delta N, S \rangle - \langle \delta X, Q \rangle\} dt \wedge du \\
&= \int_W \{\langle \dot{\eta} + \text{ad}_N \eta, S \rangle - \langle \eta' + \text{ad}_X \eta, Q \rangle\} dt \wedge du \\
&= \int_W \left\{ \left\langle \eta, \left(-\frac{\partial}{\partial t} - \text{ad}_N^* \right) S \right\rangle - \left\langle \eta, \left(-\frac{\partial}{\partial u} - \text{ad}_X^* \right) Q \right\rangle \right\} dt \wedge du \quad (5.78) \\
&\quad + \int_0^{L_0} [\langle \eta, S \rangle]_0^T du - \int_0^T [\langle \eta, Q \rangle]_0^{L_0} dt \\
&= 0
\end{aligned}$$

where we used integration-by-parts and where $\text{ad}_A^* : \mathfrak{se}(3)^* \rightarrow \mathfrak{se}(3)^*$ is the dual of the adjoint action, defined as $\langle \text{ad}_A^* B, C \rangle = -\langle B, \text{ad}_A C \rangle$. We impose that the variation vanishes at the temporal boundaries, to get

$$\mathcal{D}_t^* S = \mathcal{D}_u^* Q \quad (5.79a)$$

$$Q = 0, \quad u = 0, L_0 \quad (5.79b)$$

where

$$\mathcal{D}_t^* = \partial_t + \text{ad}_N^* \quad (5.80a)$$

$$\mathcal{D}_u^* = \partial_u + \text{ad}_X^* \quad (5.80b)$$

which was first derived in [210, 216], where Eq. 5.79b arises from imposing that $\int_0^T [\langle Q, \eta \rangle]_0^{L_0} dt = 0$ for arbitrary variations η . We call \mathcal{D}_t^* and \mathcal{D}_u^* the *dual generalised material derivatives*. As $Q = Q(X)$, Eq. 5.79b is a boundary condition on X . Equation 5.79 is the local balance law of linear and angular momentum in the absence of external forces, and together with Eq. 5.46 completely determines the constitutive kinodynamics of a Cosserat rod. These equations are first-order partial differential equations in both time t and space u . The corresponding equations of motion for a closed rod are identical, with the exception that Eq. 5.79b no longer applies.

We now evaluate Eq. 5.79 in terms of the components of S , Q , X and Y . To do this we must first compute action of the dual adjoint ad_B^* . As ad_B is a linear operator, it can be represented in matrix form as

$$[\text{ad}_A] = \begin{pmatrix} \hat{n} & \hat{a} \\ 0_{3 \times 3} & \hat{n} \end{pmatrix} \in \mathbb{R}^{6 \times 6} \quad (5.81)$$

where $A = \{\vec{a}; \hat{n}\}$ and acts as $[\text{ad}_A]\vec{B}$, where we have represented $B = \{\vec{b}; \vec{m}\}$ as vectors

$\vec{B} = (\vec{b}^T \vec{m}^T)^T \in \mathbb{R}^6$. Now we have

$$\begin{aligned}\langle B, \text{ad}_A C \rangle &= \vec{B}^T [\text{ad}_A] \vec{C} \\ &= -([\text{ad}_A^*] \vec{B})^T \vec{C} \\ &= -\vec{B}^T [\text{ad}_A^*]^T \vec{C}.\end{aligned}\tag{5.82}$$

We thus have that

$$[\text{ad}_A^*] = -[\text{ad}_A]^T = \begin{pmatrix} \hat{n} & 0_{3 \times 3} \\ \hat{a} & \hat{n} \end{pmatrix} \in \mathbb{R}^{6 \times 6}\tag{5.83}$$

where we used $\hat{a}^T = -\hat{a}$ for anti-symmetric matrices. Using Eq. 5.83 we get

$$D_t \vec{P} = D_u \vec{F}\tag{5.84a}$$

$$D_t \vec{L} = D_u \vec{M} + \vec{\theta} \times \vec{F}\tag{5.84b}$$

$$\vec{M} = 0, \quad u = 0, L_0\tag{5.84c}$$

$$\vec{F} = 0, \quad u = 0, L_0\tag{5.84d}$$

where we used the definitions of S , Q , X and Y and the covariant derivative. In the non-moving frame the equations of motion are

$$\dot{\mathbf{P}} = \mathbf{F}'\tag{5.85a}$$

$$\dot{\mathbf{L}} = \mathbf{M}' + \mathbf{r}' \times \mathbf{F}\tag{5.85b}$$

where $\mathbf{F} = \mathbf{M} = 0$ at $u = 0, L_0$, which are the classical force and moment balance equations for a Cosserat rod [1, 198]. Here we see explicitly that \mathbf{F} and \mathbf{M} is a constitutive force on the center-line and a constitutive moment on the material frame respectively, with dimensions of force and moment respectively. We can also identify \mathbf{P} as the linear momentum per unit material length of the center-line, and \mathbf{L} as the angular momentum per unit material length of the material frame. The second term in Eq. 5.85b is the moment exerted on the material frame by the force. Note that linear force balance Eq. 5.85a is the 1-dimensional analogue of the Cauchy momentum equation Eq. 4.1, where the force that arise from constitutive stresses is here \mathbf{F} . We thus see that the physical description of the Cosserat rod is consistent with classical continuum mechanics. For a discussion on how the dimensions of the kinodynamic quantities are arrived at, as well as how to nondimensionalise the dynamics, see Appendix E.

The generalised momentum can be written in a compact form as

$$\vec{S} = \mathbf{M} \vec{N}\tag{5.86}$$

where

$$\mathbf{M} = \begin{pmatrix} \rho_0 \mathbb{1}_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & \mathbb{I} \end{pmatrix} \quad (5.87)$$

is a generalised mass matrix. Similarly, in many applications the constitutive potential energy density can be written as a quadratic form

$$\mathcal{U} = \frac{1}{2}(\vec{X} - \vec{X}_0)^T \mathbf{K}(\vec{X} - \vec{X}_0) \quad (5.88)$$

where \mathbf{K} is a symmetric and positive-definite matrix, representing the elastic stiffness of the rod, and \vec{X}_0 is a given rest state. In which case the generalised stress can be written as

$$\vec{Q} = \mathbf{K}\vec{X}. \quad (5.89)$$

Henceforth we will write $S = MN$ and $Q = \mathbf{K}X$, where the action of matrices like \mathbf{M} are to be understood in the sense of Eq. 5.86.

Body forces and non-conservative dynamics

In the preceding sections we considered a stress $Q = \frac{\partial \mathcal{U}}{\partial X}$, leading to conservative dynamics. In general, the dynamics does not have to be variational (i.e. derived from a potential \mathcal{U}). When this is not the case, we say that the dynamics is *non-conservative*. Furthermore external or internal *body* forces and moments, like gravity and friction, are absent in Eq. 5.79.

The following is a generalisation of the derivations in [214, 215, 217] for continuum systems. We write the continuum analogue of the *integral Lagrange-d'Alembert* principle as [42, 212]

$$\delta \int_W \mathcal{L}(d\Phi) dt \wedge du + \int_W (\mathbf{T}(\Phi, d\Phi, \dots))(\delta\Phi) dt \wedge du = 0 \quad (5.90)$$

where $\mathbf{T}(\Phi, d\Phi, \dots) \in T^*SE(3)$ is a covector field we call the *generalised body force* and can be an arbitrary function of Φ and its derivatives, and where it is acting on the variational test function $\delta\Phi$. Henceforth we will suppress its arguments and write $(\mathbf{T}(\Phi, d\Phi, \dots))(\delta\Phi) = \mathbf{T}(\delta\Phi)$. Analogous to previous section, we presume that the Lagrangian density can be reformulated to a reduced form $\ell(\xi)$. Compare with Eq. 5.27. Note that the way \mathbf{T} appears in the integral, as a density over the kinematic base space W , shows that it is indeed a generalised *body* force. Eq. 5.90 is a generalisation of Hamilton's principle Eq. 5.74 that also incorporates body forces.

By the Euler–Poincaré theorem [215], the first term in Eq. 5.90 evaluates as previously

to Eq. 5.78. The second term can be written as

$$\begin{aligned}\int_W \mathbf{T}(\delta\Phi) dt \wedge du &= \int_W \mathbf{T}(\Phi\eta) dt \wedge du \\ &= \int_W \mathbf{T}(L_\Phi\eta) dt \wedge du \\ &= \int_W \langle \eta, T \rangle dt \wedge du\end{aligned}\tag{5.91}$$

where we used $\eta = \Phi^{-1}\delta\Phi$, and where $T = L_\Phi^*\mathbf{T}$ is the generalised body force mapped to the dual Lie algebra and is thus a map $T : W \rightarrow \mathfrak{se}(3)^*$, and $L_\Phi^* : T_\Phi^*SE(3) \rightarrow \mathfrak{se}(3)^*$ is a mapping from the cotangent bundle to the dual Lie algebra defined as $\mathbf{T}(L_\Phi\eta) = \langle L_\Phi^*\mathbf{T}, \eta \rangle$. The resulting dynamical equations of motion, which now include arbitrary body forces and moments, are

$$\mathcal{D}_t^*S = \mathcal{D}_u^*Q + T\tag{5.92a}$$

$$Q = 0, \quad u = 0, L_0.\tag{5.92b}$$

The analogous equation for non-continuum systems can be found Eq. 4 in [202] and Eq. 6 in [215]. Expressed in non-Lie algebraic terms, by setting $\vec{T} = \{\vec{f}; \vec{m}\}^*$, the equations of motion in the moving frame are

$$D_t \vec{P} = D_u \vec{F} + \vec{f}\tag{5.93a}$$

$$D_t \vec{L} = D_u \vec{M} + \vec{\theta} \times \vec{F} + \vec{m}\tag{5.93b}$$

where \vec{f} and \vec{m} are the body force and moment respectively, with units of force and moment per unit material length respectively, and $\vec{F} = \vec{M} = 0$ at $u \in 0, L_0$ as before. Finally, in the non-moving frame, we have

$$\dot{\mathbf{P}} = \mathbf{F}' + \mathbf{f}\tag{5.94a}$$

$$\dot{\mathbf{L}} = \mathbf{M}' + \mathbf{r}' \times \mathbf{F} + \mathbf{m}\tag{5.94b}$$

where $\mathbf{F} = \mathbf{M} = 0$ at $u = 0, L_0$. Note that \mathbf{f} and \mathbf{m} (and \vec{f} and \vec{m}) have dimension force per unit material length and moment per unit material length respectively.

Although technical, the derivation in Eq. 5.91 shows in what form the generalised body force appears in the equations of motion. In practice however, we do not need to evaluate $L_\Phi^*\mathbf{T} \in \mathfrak{se}(3)^*$, but can instead find the expression for the body forces and moments by first defining them in the non-moving frame. For example, to incorporate gravity we would have a body force $\mathbf{f} = -\rho_0 g \mathbf{d}_3$, where g is here the gravitational constant. In the moving frame, the force is then $\vec{f} = E^T \mathbf{f}$. We see that the static force \mathbf{f} becomes state-dependent

in the moving frame. In this context \vec{f} is referred to as an *advection force*, and is treated in detail in [212, 218, 219].

As the example of gravity highlights, in general body forces and moments can be functions of the Lie group-level configuration Φ , or indeed functions of space-time (t, u) or the Lie algebra-valued X and N , and any of their derivatives. In this general case the dynamics are thus no longer defined purely on the Lie algebra. For simulations, this mean that Φ needs to be reconstructed from ξ for every time t , in order to evaluate the dynamics. In Sec. 5.2.1 we showed how to reconstruct Φ from ξ .

Finally, we now consider non-variational constitutive forces. For a point particle, the Euler-Lagrange equations can be shown to be a special case of D'Alembert's principle where the applied force derives from a potential energy function. Consider a point particle with configuration space Q , and a Lagrangian density $\mathcal{L} : Q \times TQ \rightarrow \mathbb{R}$ given by $\mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}) = \mathcal{K}(\dot{\mathbf{q}}) - \mathcal{U}(\mathbf{q})$, where \mathbf{q} are coordinates on Q . Then from Hamilton's principle $\delta \int_0^T \mathcal{L} = 0$ we find

$$\int_0^T \left(\frac{\partial \mathcal{K}}{\partial \dot{\mathbf{q}}} \cdot \delta \dot{\mathbf{q}} - \frac{\partial \mathcal{U}}{\partial \mathbf{q}} \cdot \delta \mathbf{q} \right) dt = 0. \quad (5.95)$$

We can identify $\mathbf{P} = \frac{\partial \mathcal{K}}{\partial \dot{\mathbf{q}}}$ and $\mathbf{F} = \frac{\partial \mathcal{U}}{\partial \mathbf{q}}$ as the inertial and external forces respectively, which we also identify as covectors $\mathbf{P}, \mathbf{F} \in T^*Q$ as they contract with tangent vectors. Now if we promote F_i to be an arbitrary non-variational force, then we have

$$\int_0^T (\mathbf{P} \cdot \delta \dot{\mathbf{q}} - \mathbf{F} \cdot \delta \mathbf{q}) dt = 0. \quad (5.96)$$

which is D'Alembert's principle in integral form for a point particle.

By analogy to the previous example, we construct D'Alembert's principle for the constitutive dynamics of a Cosserat rod, which reduce to the Euler-Poincaré equations when the forces and moments are conservative. As shown in Eq. 5.78, the variation of $\int_W \ell dt \wedge du$ leads to

$$\int_W \{ \langle \delta N, S \rangle - \langle \delta X, Q \rangle \} dt \wedge du = 0, \quad (5.97)$$

where previously $Q = \frac{\partial U}{\partial X}$ was variational. However if we let Q to be non-variational in general, we can take Eq. 5.97 to be a D'Alembert's principle for constitutive mechanics in the absence of body forces. Thus we present a final generalisation of the integral Lagrange-d'Alembert principle

$$\int_W \{ \langle \delta N, S \rangle - \langle \delta X, Q \rangle \} dt \wedge du + \int_W \langle \eta, T \rangle dt \wedge du = 0 \quad (5.98)$$

which incorporates both body and non-conservative forces and moments. The resulting equations of motion are identical to Eq. 5.92, but where Q is now an arbitrary function of X .

Equation 5.98 should be compared with Eq. 5.28, which is the corresponding principle for rigid bodies. As rigid bodies are in-effect modelled as oriented point particles, they suffer no internal stresses. There is therefore no notion of a constitutive dynamics for a rigid body, as there are for a continuum bodies. For the latter, the first term in Eq. 5.28 is now instead the first term in Eq. 5.98, which accommodates for general non-conservative internal stresses.

5.2.3 Summary and discussion

Here we give a brief overview and summary of the steps taken to derive the kinodynamic equations of motion of the Cosserat rod. In Eq. 5.31 we showed that the configuration space of the Cosserat rod can be seen as the Lie group of special Euclidean transformations $SE(3)$, and its spatio-temporal configuration can thus be seen as a group-valued function $\Phi(t, u) \in SE(3)$, where $t \in [0, T]$ and $u \in [0, L_0]$. We referred to $M = [0, L_0]$ as the material base space, and $W = [0, T] \times M$ the kinematic base space. Through the Lie group-Lie algebra correspondence, Eq. 5.35, the configuration of the rod can additionally be described as a Lie algebra-valued function $\xi(t, u) \in \mathfrak{se}(3)$. The kinematic equations of motion Eq. 5.46, formulated in the Lie algebra, can then be derived by imposing the integrability condition Eq. 5.36.

Dynamically, we modelled the Cosserat rod as a kinematic approximation of a slender three-dimensional rod. We wrote down the general Lagrangian density for the constitutive mechanics of the latter in Eq. 5.62, and then proceeded to derive the corresponding reduced Lagrangian density defined on the Lie algebra Eq. 5.67. By performing Lie algebraic variations of the reduced Lagrangian density, we arrived at the Euler-Poincaré for the constitutive dynamics of the Cosserat rod Eq. 5.79.

Finally, we derived a generalised integral Lagrange-d'Alembert principle Eq. 5.98, that allows for both non-variational (non-conservative) constitutive forces and moments, as well as body forces and moments.

The full set of kinodynamical equations of motion for the Cosserat rod can thus be written as

$$\dot{X} = \mathcal{D}_u N \quad (5.99a)$$

$$\mathcal{D}_t^* S = \mathcal{D}_u^* Q + T \quad (5.99b)$$

$$Q = 0, \quad u = 0, L_0, \quad (5.99c)$$

and the equations close using $P = MN$. Using $X = \{\vec{\theta}; \vec{\pi}\}$, $N = \{\vec{V}; \vec{\Omega}\}$, $S = \{\vec{P}; \vec{L}\}^*$,

$Q = \{\vec{F}; \vec{M}\}^*$ and $T = \{\vec{f}; \vec{m}\}^*$ the equations of motion in the moving frame become

$$D_t \vec{\theta} = D_u \vec{V} \quad (5.100a)$$

$$\partial_t \vec{\pi} = D_u \vec{\Omega} \quad (5.100b)$$

$$D_t \vec{P} = D_u \vec{F} + \vec{f} \quad (5.100c)$$

$$D_t \vec{L} = D_u \vec{M} + \vec{\theta} \times \vec{F} + \vec{m} \quad (5.100d)$$

with $\vec{F} = \vec{M} = 0$ at $u = 0, L_0$, and the equations close using $\vec{P} = \rho_0 \vec{V}$ and $\vec{L} = \mathbb{I} \vec{\Omega}$, and D_u and D_t are defined in Eq. 5.44.

The equations of motion that would have resulted from varying the Lagrangian density $\mathcal{L}(\Phi, d\Phi)$ in Eq. 5.74, defined on the Lie group, would have resulted in equations of motion defined on the *non-linear* space $SE(3) \times TSE(3)$ and $SE(3) \times T^*SE(3)$. In contrast, varying the reduced Lagrangian density Eq. 5.73 results in equations of motion Eq. 5.99 defined on $W \times \mathfrak{se}(3)$ and $W \times \mathfrak{se}(3)^*$, which are linear spaces². This drastic simplification of the dynamics is inherently made possible due to the properties of a Lie group G , allowing for the trivialisation $G \times TG \cong G \times \mathfrak{g}$.

When the dynamics is constitutive then it is notable that the above equations of motion are entirely formulated within the Lie algebra. That is, the dynamical variables are the Lie algebra-valued X and N , and all forces and moments are expressed as functions of them. The invariance of the constitutive dynamics under rigid body rotations and translations is inherent in Eq. 5.99, as all quantities in the equations are in themselves invariant.

In the next section we will consider kinematically constrained Cosserat rods, and in particular the *filament*, which has as kinematic configuration space the homogeneous space $SE(3)/SO(3)$. The results of this and the next section will be generalised in Ch. 6 to systems where the kinematic configuration space is a general Lie group or homogeneous space, and the material base space is a general manifold.

We conclude this section with the remark that it is possible to view the mathematical formalism that we have developed from a field-theoretic lens. The spatio-temporal configuration Φ can be seen as a $SE(3)$ -valued field defined over the base space $M = [0, L_0]$. The kinodynamic equations of motion are then the field equations, defined in terms of the Lie algebra.

5.3 Applications

Here we give a few examples of Cosserat rod models as applied in physics. In Sec. 5.3.1 we derive, from first principles, the constitutive dynamics of a coarse-grained slender rod. Using the results of the previous section, we then approximate the kinodynamics of the

²Note that this is not strictly-speaking true in general, if T is a function of Φ

slender rod using the Cosserat rod. In Sec. 5.3.2 we discuss Cosserat rods under the influence frictional forces. We formulate the equations of motion in both the underdamped and overdamped limits. Finally, in Sec. 5.3.3 we derive the system of equations that describe an inextensible Cosserat rod, in the overdamped limit.

5.3.1 Hyperelastic slender rod

In many applications the conservative part of the constitutive dynamics arise from a potential of the form Eq. 5.88 [220, 221, 222, 223], where the potential \mathcal{U} is a quadratic form in X , with coefficients given by the symmetric and positive-definite stiffness matrix $K \in \mathbb{R}^{6 \times 6}$. Here we derive the form of K from first principles, given constitutive assumptions on the material that the rod is composed of. Specifically, we use the Cosserat to approximate the continuum mechanics of a slender rod of hyper-elastic *Saint Venant-Kirchoff* material [224].

We consider a straight rod $\mathcal{M} = \mathcal{D} \times [0, L_0]$ at rest, where \mathcal{D} is a disc of arbitrary shape, and L_0 is the length of the rod. As before, \mathbf{X} is the material coordinate on \mathcal{M} , and we align $X_1 \in [0, L_0]$ along the length of the rod, and $(X_2, X_3) \in \mathcal{D}$. We will use $u = X_1$ and X_1 interchangeably. Deformations of \mathcal{M} are given by $\mathbf{x}(\mathbf{X})$, which we will refer to as the *deformed coordinate*. The material and deformed coordinates can be related as

$$d\mathbf{x} = \mathcal{P}d\mathbf{X} \quad (5.101)$$

where $\mathcal{P} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}}$ is here the *deformation gradient tensor*, and we write its components in the moving frame $\mathcal{P}_{ij} = \mathbf{e}_i \cdot (\frac{\partial \mathbf{x}}{\partial \mathbf{X}} \mathbf{e}_j)$. Following [225], we consider the change in infinitesimal distances between two material points \mathbf{X}_1 and \mathbf{X}_2 in the rod after a deformation. In the undeformed state, the distance between \mathbf{X}_1 and \mathbf{X}_2 is $d\ell^2 = |d\mathbf{X}|^2$. After deformation we have

$$\begin{aligned} d\ell'^2 &= |d\mathbf{x}|^2 = (\mathcal{P}d\mathbf{X})^T(\mathcal{P}d\mathbf{X}) \\ &= d\mathbf{X}^T \mathcal{P}^T \mathcal{P} d\mathbf{X} \\ &= d\ell^2 + 2d\mathbf{X}^T E d\mathbf{X} \\ &= d\mathbf{X}^T (I + 2E) d\mathbf{X} \end{aligned} \quad (5.102)$$

where

$$E = \frac{1}{2} \left(\frac{\partial \mathbf{u}}{\partial \mathbf{X}} + \left(\frac{\partial \mathbf{u}}{\partial \mathbf{X}} \right)^T + \left(\frac{\partial \mathbf{u}}{\partial \mathbf{X}} \right)^T \frac{\partial \mathbf{u}}{\partial \mathbf{X}} \right) \quad (5.103)$$

is the *Lagrangian strain tensor* and $\mathbf{u}(\mathbf{X}) = \mathbf{x} - \mathbf{X}$ is the *deformation vector*. For small deformations, we can approximate Eq. 5.108 as $E \approx \frac{1}{2} \left(\frac{\partial \mathbf{u}}{\partial \mathbf{X}} + \left(\frac{\partial \mathbf{u}}{\partial \mathbf{X}} \right)^T \right)$, in which case we can express it as

$$E = \frac{1}{2} (\mathcal{P} + \mathcal{P}^T - 2\mathbb{I}). \quad (5.104)$$

The Lagrangian strain tensor can be seen as giving the relative deformations along the principle axes of a material, where the latter are defined as the eigenvectors E . The elastic energy density of the material in a deformed state is a function of E , and we assume it to be a quadratic form

$$W(E) = \frac{1}{2}C_{ijkl}E_{ij}E_{kl} \quad (5.105)$$

where C_{ijkl} is a fourth-order tensor of elastic constants, which is known as the Saint Venant-Kirchoff constitutive assumption [224], and materials that are well-approximated under this assumption are known as *Saint Venant-Kirchoff materials*. If we assume that C is isotropic, in which case it has the form

$$C_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu_1\delta_{ik}\delta_{jl} + \mu_2\delta_{il}\delta_{jk} \quad (5.106)$$

then Eq. 5.105 can be shown to reduce to

$$W(E) = \mu\text{Tr}(E^2) + \frac{\lambda}{2}\text{Tr}(E)^2 \quad (5.107)$$

where λ and $\mu = \mu_1 + \mu_2$ are known as the *Lamé constants*.

We will now proceed to derive the potential energy per unit material length of a Cosserat rod under the constitutive assumption of Eq. 5.107. We impose the kinematic assumption Eq. 5.61, which we write again here as

$$\mathbf{x}(\mathbf{X}) = \mathbf{r}(X_1) + \mathbf{e}_\xi X_\xi, \quad \xi = 2, 3 \quad (5.108)$$

which constrains deformations to preserve the area and shape of the cross-section at each X_1 . Since

$$\frac{\partial \mathbf{x}}{\partial X_1} = \theta_i \mathbf{e}_i + \mathbf{e}_i \hat{\pi}_{i\gamma} X_\gamma, \quad \gamma = 2, 3 \quad (5.109)$$

where we used Eq. 5.40, we have

$$\begin{aligned} \mathcal{P}_{i1} &= \theta_i + \pi_{i\gamma} X_\gamma \\ \mathcal{P}_{i\gamma} &= \delta_{i\gamma}. \end{aligned} \quad (5.110)$$

The Lagrangian strain tensor E can then be computed using Eq. 5.104, and since Eq. 5.110 is expressed entirely in terms of $X = \{\vec{\theta}; \vec{\pi}\}$, we can express the energy volume density W as a function of X as well. To find the energy density per unit-length $\mathcal{U}(\vec{\theta}, \vec{\pi})$ of the rod, we must integrate out the cross-section

$$\mathcal{U}(\vec{\theta}, \vec{\pi}) = \int_{\mathcal{D}} W(E) dX_2 dX_3. \quad (5.111)$$

For simplicity, we assume a circular cross-section, for which we have that

$$\begin{aligned}
\int_{\mathcal{D}} dA &= \pi R^2 \\
\int_{\mathcal{D}} X_\alpha dA &= 0 \\
\int_{\mathcal{D}} X_\alpha X_\beta dA &= \frac{\pi R^4}{4} \delta_{\alpha\beta} \\
\int_{\mathcal{D}} X_\alpha X_\beta X_\gamma dA &= 0 \\
\int_{\mathcal{D}} X_\alpha X_\beta X_\gamma X_\delta dA &= \begin{cases} \frac{\pi R^6}{8}, & \alpha = \beta = \gamma = \delta \\ \frac{\pi R^6}{24}, & \alpha = \beta \text{ and } \gamma = \delta \\ \frac{\pi R^6}{24}, & \alpha = \gamma \text{ and } \gamma = \beta \end{cases}
\end{aligned} \tag{5.112}$$

where $\alpha, \beta, \gamma, \delta = 2, 3$ and R is the radius of the rod. From Eq. 5.112, Eq. 5.112, Eq. 5.104 and Eq. 5.107, the potential energy per unit material length evaluates to

$$\mathcal{U}(\vec{\theta}, \vec{\pi}) = \frac{k_1}{2}(\theta_1 - 1)^2 + \frac{k_2}{2}(\theta_2^2 + \theta_3^2) + \frac{\epsilon_1}{2}\hat{\pi}_1^2 + \frac{\epsilon_2}{2}(\hat{\pi}_2^2 + \hat{\pi}_3^2) \tag{5.113}$$

where $k_1 = \pi R^2(\lambda + 2\mu)$, $k_2 = \frac{\pi R^2}{2}\lambda$, $\epsilon_1 = \frac{\pi R^4}{4}\lambda$ and $\epsilon_2 = \frac{\pi R^4}{4}(\lambda + 2\mu)$. Equation 5.113 is consistent with the quadratic Cosserat free energies found in the literature [196, 210, 226]. For a non-circular cross-section we will in general have a potential energy of the form

$$\mathcal{U}(\vec{\theta}, \vec{\pi}) = \frac{k_1}{2}(\theta_1 - 1)^2 + \frac{k_2}{2}\theta_2^2 + \frac{k_3}{2}\theta_3^2 + \frac{\epsilon_1}{2}\hat{\pi}_1^2 + \frac{\epsilon_2}{2}\hat{\pi}_2^2 + \frac{\epsilon_3}{2}\hat{\pi}_3^2. \tag{5.114}$$

We can interpret the form of the potential energy by recalling how $\vec{\theta}$ and $\vec{\pi}$ encodes the various kinematic deformations of the Cosserat rod, described in Sec. 5.2.1. The first three terms in Eq. 5.114 penalises both shearing of the material frame, as well as extension of the center-line. The π_1 term penalises twisting of the material frame along \mathbf{e}_1 , and π_2 and π_3 penalises the rotation of the material frame along \mathbf{e}_2 and \mathbf{e}_3 respectively.

The stiffness matrix of a hyper-elastic Cosserat rod can thus be written as

$$\mathbf{K} = \begin{pmatrix} K^{(1)} & 0_{3 \times 3} \\ 0_{3 \times 3} & K^{(2)} \end{pmatrix} \tag{5.115}$$

where $K^{(1)} = \text{diag}\{k_1, k_2, k_3\}$ and $K^{(2)} = \text{diag}\{\epsilon_1, \epsilon_2, \epsilon_3\}$, with rest state

$$\vec{\theta}_0 = (1 \ 0 \ 0)^T \tag{5.116a}$$

$$\vec{\pi}_0 = (0 \ 0 \ 0)^T. \tag{5.116b}$$

The resulting constitutive force and moments are

$$\vec{F} = \frac{\partial \mathcal{U}}{\partial \vec{\theta}} = K^{(1)}(\vec{\theta} - \vec{\theta}_0) \quad (5.117a)$$

$$\vec{M} = \frac{\partial \mathcal{U}}{\partial \vec{\pi}} = K^{(2)}\vec{\pi}. \quad (5.117b)$$

We see that in this constitutive model, the force \vec{F} acts on the center-line such as to align with its material frame. In other words, if the material frame is shearing, then the center-line will bend such as to be tangent to \mathbf{e}_1 . If the material frame is rotating along u , which corresponds to a non-zero $\vec{\pi}$, then the moment \vec{M} acts so as to render it non-rotating. We thus see that a shearing material frame *induces* bend. This effect happens despite the fact that $\vec{\theta}$ and $\vec{\pi}$ do not couple in \mathcal{U} , which inherently due to the geometric non-linearities in the kinematics of the Cosserat rod. Conversely, a bending center-line which is aligned with the material frame will induce shear. Note that there is no direct energetic cost for the bending of the center-line itself (as \mathcal{U} is not a function of the scalar curvature κ), however the energetic cost of shearing, combined with the energetic cost of a rotating material frame, conspire together so as to straighten segments of the rod that bend.

5.3.2 Dissipative and overdamped dynamics

Equation 5.93 allows us to model general dissipative dynamics. We consider the case of a Cosserat rod that suffers both internal and external friction

$$\dot{X} = \mathcal{D}_u X \quad (5.118a)$$

$$\mathcal{D}_t^* S = \mathcal{D}_u^* Q - \mathsf{H} \dot{X} - \mathsf{G} N \quad (5.118b)$$

$$Q = 0, \quad u = 0, L_0, \quad (5.118c)$$

where we have set $T = -\mathsf{H} \dot{X} - \mathsf{G} N$, and where H and G are positive-definite matrices of damping coefficients of the internal and external friction respectively. Note that $\mathsf{H} \dot{X}$ in Eq. 5.118b is evaluated by substituting in Eq. 5.118a. If let the frictional matrices be block-diagonal in the translational and angular components as

$$\mathsf{H} = \begin{pmatrix} \gamma_T^{\text{in}} & 0_{3 \times 3} \\ 0_{3 \times 3} & \gamma_R^{\text{in}} \end{pmatrix} \quad (5.119a)$$

$$\mathsf{G} = \begin{pmatrix} \gamma_T & 0_{3 \times 3} \\ 0_{3 \times 3} & \gamma_R \end{pmatrix} \quad (5.119b)$$

where γ_T^{in} , γ_R^{in} , γ_T and γ_R are positive-definite 3×3 -matrices, we can write the equations of motion as

$$D_t \vec{\theta} = D_u \vec{V} \quad (5.120\text{a})$$

$$\partial_t \vec{\pi} = D_u \vec{\Omega} \quad (5.120\text{b})$$

$$D_t \vec{P} = D_u \vec{F} - \gamma_T^{\text{in}} \dot{\vec{\theta}} - \gamma_T \vec{V} \quad (5.120\text{c})$$

$$D_t \vec{L} = D_u \vec{M} + \vec{\theta} \times \vec{F} - \gamma_R^{\text{in}} \dot{\vec{\pi}} - \gamma_R \vec{\Omega} \quad (5.120\text{d})$$

where again $\gamma_T^{\text{in}} \dot{\vec{\theta}}$ and $\gamma_R^{\text{in}} \dot{\vec{\pi}}$ can be evaluated using the kinematic equations of motion, and $\vec{F} = \vec{M} = 0$, $u = 0, L_0$.

We can consider a limit where the dissipative forces dominate over the inertial forces (which correspond to the left-hand side of Eq. 5.118b). In this limit, we set $S \approx 0$, or equivalently $\vec{P} \approx 0$ and $\vec{L} = \mathbb{I}\vec{\Omega} \approx 0$. The resulting equations are *overdamped* equations of motion

$$\dot{\vec{X}} = \mathcal{D}_u N \quad (5.121\text{a})$$

$$\mathbf{G}N = \mathcal{D}_u^* Q + \mathbf{H}\dot{\vec{X}} \quad (5.121\text{b})$$

$$Q = 0, \quad u = 0, L_0. \quad (5.121\text{c})$$

where we must solve for N in Eq. 5.121b to compute Eq. 5.121a. In terms of the translational and angular velocities, the equations of motion are

$$D_t \vec{\theta} = D_u \vec{V} \quad (5.122\text{a})$$

$$\partial_t \vec{\pi} = D_u \vec{\Omega} \quad (5.122\text{b})$$

$$\gamma_T \vec{V} = D_u \vec{F} - \gamma_T^{\text{in}} \dot{\vec{\theta}} \quad (5.122\text{c})$$

$$\gamma_R \vec{\Omega} = D_u \vec{M} + \vec{\theta} \times \vec{F} - \gamma_R^{\text{in}} \dot{\vec{\pi}} \quad (5.122\text{d})$$

where again we must solve for \vec{V} and $\vec{\Omega}$ in the latter two equations and substitute them into the kinematic equations of motion. We note that the overdamped equations of motion are first-order in time t and second-order in space u (as opposed to first-order in the underdamped case). Furthermore, the only degree of freedom is here $X = \{\vec{\theta}; \vec{\Omega}\}$, and so the dynamics are 6-dimensional as opposed to 12-dimensional in the underdamped case.

The presence of the internal friction $\mathbf{H}\dot{\vec{X}}$ significantly complicates the evaluation of the overdamped dynamics. Substituting Eq. 5.121a into Eq. 5.121b we get a self-consistency relation for N

$$\mathbf{G}N = \mathcal{D}_u^* Q + \mathbf{H}\mathcal{D}_u N \quad (5.123)$$

and formally we can write N as

$$N = \mathcal{A}^{-1} \mathcal{D}_u^* Q \quad (5.124)$$

where $\mathcal{A} = \mathbf{G} - \mathbf{H}(\partial_u + \text{ad}_X)$. In simulations, which invariably employ some discretisation scheme, \mathcal{A} becomes a matrix that can be inverted numerically. If, however, we neglect internal dissipation we have $N = \mathbf{G}^{-1}(\partial_u - \text{ad}_X^*)Q$, and we can write the equations of motion in a compact form as

$$\dot{\mathbf{X}} = \mathcal{D}_u [\mathbf{G}^{-1} \mathcal{D}_u^* Q], \quad (5.125)$$

or equivalently

$$D_t \vec{\theta} = D_u [\mu_T D_u \vec{F}] \quad (5.126a)$$

$$\partial_t \vec{\pi} = D_u [\mu_R D_u \vec{M}] \quad (5.126b)$$

where $\mu_{T,R} = \gamma_{T,R}^{-1}$, which would be the overdamped equations of motion of a Cosserat rod immersed in a viscous medium.

5.3.3 Overdamped and inextensible Cosserat rods

We will now consider applying the kinematic constraint of *inextensibility* to a Cosserat rod in the overdamped limit. We will assume μ_T is a diagonal matrix and can thus be treated as a scalar constant. As in [178, 227, 228], we enforce inextensibility by adding a constitutive tension \vec{T} along the direction tangent to the center-line. To simplify our calculations, we will decompose the constitutive force as $\vec{F}_{\text{tot}} = \vec{F} + \vec{T}$. As $\boldsymbol{\theta}$ is a vector pointing in the tangential direction, we will also let $\vec{T} = T \vec{\theta}$, where T is a scalar function. Now, T must now be chosen such that the constraint $\mathbf{r}' \cdot \mathbf{r}' = \boldsymbol{\theta} \cdot \boldsymbol{\theta} = 1$ is enforced. Now, we have

$$\vec{\theta}^T D_t \vec{\theta} = \vec{\theta}^T \dot{\vec{\theta}} = \frac{1}{2} \partial_t (\vec{\theta}^T \vec{\theta}) = \frac{1}{2} \partial_t (\boldsymbol{\theta} \cdot \boldsymbol{\theta}). \quad (5.127)$$

where we used the anti-symmetry of $\hat{\Omega}$. From Eq. 5.126a we therefore find that $\vec{\theta}^T D_u^2 (\vec{F} + T \vec{\theta}) = 0$. By expanding out the derivative, and using $\partial_t (\vec{\theta}^T \vec{\theta}) = \partial_u (\vec{\theta}^T \vec{\theta}) = 0$ and $\vec{\theta}^T \vec{\theta} = 1$ we find

$$(\partial_u^2 - H)T = -\vec{\theta}^T D_u^2 \vec{F} \quad (5.128)$$

where

$$H = -\vec{\theta}^T (\vec{\theta}'') + 2\vec{\theta}^T \hat{\pi} (\vec{\theta}') - \vec{\theta}^T \hat{\pi} \hat{\pi} \vec{\theta} \quad (5.129)$$

In numerical simulations, the non-linear 2nd order ODE Eq. 5.128 would have to be solved at each time-step in order to compute the tensional force.

We now consider the case of a twisting filament, as defined in Sec. 5.4.1, for which the

$$\mathcal{F}(u, t) \in SE(3)$$

$$\mathcal{F}(u, t) \in SE(3)/SO(3) \cong \mathbb{E}^3$$

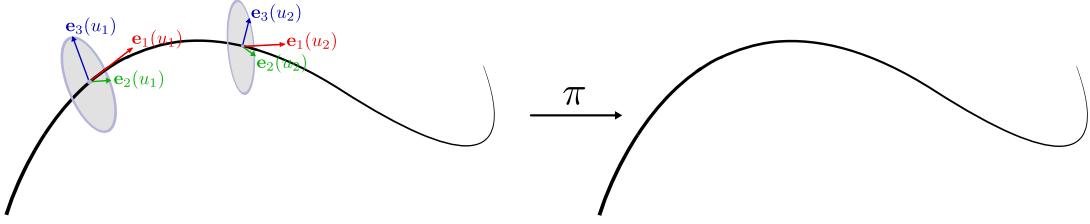


Figure 5.4: (Left) A Cosserat rod, with configuration space $SE(3)$. (Right) A filament, with configuration space in $SE(3)/SO(3) \cong \mathbb{E}^3$.

material frame is adapted to the center-line. Then, as $\vec{\theta} = (1 \ 0 \ 0)^T$, Eq. 5.128 simplifies to

$$(\partial_u^2 - (\pi_2^2 + \pi_3^2)) T = -(D_u^2 \vec{F})_1. \quad (5.130)$$

5.4 Mechanics of kinematically constrained Cosserat rods

Thus far we have considered the material frame E and the center-line \mathbf{r} as independent degrees of freedom, modelling a rod that can shear and twist in addition to bending and extending. In many applications it suffices to consider pure center-line dynamics where the cross-section does not shear [178, 228, 229, 230, 231, 232]. We shall refer to these systems as *filaments*, and can be seen as a Cosserat rod model with a kinematically constrained material frame. As we will show, kinematic constraints can often be implemented by first identifying an appropriate Lie subgroup $H \subset SE(3)$ such that the homogeneous space $SE(3)/H$ is the kinematically constrained configuration space, and then by making a ‘gauge choice’ $h(u) \in H$ which fixes a parametrisation of $SE(3)/H$. In addition to the filament, we will also consider the case of a *twisting filament*, where the cross-section is allowed to twist around the center-line [177, 197, 227].

5.4.1 Kinematic constraints and gauge freedoms

The configuration space of a filament can be found by ‘rotating away’ the material frame $E(u)$ at each u . Therefore the appropriate Lie subgroup is $H = SO(3)$, and the configuration space is $SE(3)/SO(3) \cong \mathbb{E}^3$, where \mathbb{E}^3 is the three-dimensional Euclidean space. Conceptually, that $SE(3)/SO(3)$ represents the configuration space of the filament can be understood by recalling the definition of a homogeneous space. For any Lie group G and Lie subgroup $H \subset G$ then G/H is the set

$$G/H = \{gH : g \in G\} \quad (5.131)$$

where $gH = \{gh : h \in G\}$ is the left coset. Now consider the case when $G = SE(3)$ and $H = SO(3)$. Let $g_1 = (\mathbf{r}, E_1)$ and $g_2 = (\mathbf{r}, E_2)$ denote two cross-sections that are located at the same point \mathbf{r} , but have two different material frames E_1 and E_2 , and furthermore we can trivially identify these as elements $g_1, g_2 \in SE(3)$. As E_1 is related to E_2 by a rotation in $SO(3)$, we have that $g_1H = g_2H$. In other words, in the homogeneous space $SE(3)/SO(3)$, g_1 and g_2 belong to the same element $g_1H = g_2H \in SE(3)/SO(3)$. Therefore, a filament can mathematically be described as a Cosserat rod where we have, for each u , identified all material frames $E(u)$ at $\mathbf{r}(u)$ as belonging to the same equivalence class. See Fig. 5.4 for an illustration.

Now let $\Phi : W \rightarrow SE(3)/SO(3)$ be the spatio-temporal configuration of the filament. As discussed in Sec. 4.2.6, we can write Φ as

$$\Phi = \pi \circ \tilde{\Phi} \quad (5.132)$$

where \circ denotes compositions of maps and $\pi : SE(3) \rightarrow SE(3)/SO(3)$ relates elements in $SE(3)$ to elements in the homogeneous space and can be defined as

$$\pi((\mathbf{r}; E)) = \mathbf{r}, \quad (5.133)$$

and $\tilde{\Phi} : W \rightarrow SE(3)$ is a *framing*, or an *adapted frame field*, of Φ [200], and is a Lie group-valued function on W that satisfies Eq. 5.132.

We can map any framing $\tilde{\Phi}$ to a spatio-temporal filament configuration Φ using Eq. 5.132. Note however that this map is not injective. Let $h : W \rightarrow SO(3)$ be an arbitrary $SO(3)$ -valued smooth function and let $\tilde{\Phi}_1 : W \rightarrow SE(3)$ and $\tilde{\Phi}_2 : W \rightarrow SE(3)$, then if $\tilde{\Phi}_1(t, u) = h(t, u)\tilde{\Phi}_2(t, u)$ maps to the same Φ . Any given choice of $\tilde{\Phi}$ is called a *framing* of Φ [200]. There is thus a *gauge freedom* in the choice of $\tilde{\Phi}$.

In principle, the kinodynamical equations for a Cosserat rod Eq. 5.99 apply to the filament as well. We could initialise Eq. 5.99 with some $\tilde{\Phi}(u, 0)$ that maps to the true initial condition $\Phi(u, 0)$, then simulate the equations of motion to find $\tilde{\Phi}(t, u)$, and then finally use Eq. 5.132 to find $\Phi(t, u)$. In this case, it is important that all constitutive and body forces and moments only couple to the center-line \mathbf{r} , and not the material frame E . For example, for conservative constitutive dynamics we could have a potential energy density $\mathcal{U} \propto \kappa^2$, where κ is the scalar curvature of the center-line defined in Eq. 5.52.

However, the above would mean simulating equations of motion of higher-dimensionality than the system in question, where the extra degrees of freedom (the material frame) are non-physical. In the following we will instead choose a particular *gauge* and form an injective map between $\tilde{\Phi}$ and Φ . In other words, by choosing a gauge we choose a particular prescription for how to frame a given space-curve. Beyond the fact that the resulting dynamics will be defined directly on the lower dimensional $SE(3)/SO(3)$ rather than

$SE(3)$, it is also often easier to formulate constitutive dynamics given a particular framing, rather than formulating it in terms of the degrees of freedom of a general Cosserat rod. The programmatic construction of such gauges is better known as the *theory of moving frames* [207, 233]. We note that the terminological use of ‘gauge’, although well-motivated in this context, is not standard in the theory of moving frames.

All gauge choices are ultimately mathematically equivalent, however some choices are more practically useful than others. Our procedure to construct a gauge will be to write the material frame $E = (\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3)$ as a function of the center-line \mathbf{r} . Thus one-to-one map between Φ and $\tilde{\Phi}$ is formed. As a first step towards finding an appropriate gauge, we set

$$\mathbf{e}_1 = \frac{\mathbf{r}'}{h} \quad (5.134)$$

where $h = |\mathbf{r}'|$, such that \mathbf{e}_1 is a unit-vector tangent to \mathbf{r} at u . This means that given \mathbf{e}_1 we can write the center-line as

$$\mathbf{r}(t, u) = \mathbf{r}(0, t) + \int_0^u h\mathbf{e}_1 du. \quad (5.135)$$

There is now a remaining $SO(2)$ gauge freedom in specifying the material frame vectors \mathbf{e}_2 and \mathbf{e}_3 . In the gauges found in the literature [200, 234, 235, 236, 237], most abide Eq. 5.134, and are thus all related by an $SO(2)$ gauge transformation of \mathbf{e}_2 and \mathbf{e}_3 .

The Frenet-Serret gauge

In the *Frenet-Serret* gauge [238, 239], we set

$$\mathbf{e}_2 = \frac{\mathbf{e}'_1}{\kappa} \quad (5.136)$$

where $\kappa = |\mathbf{e}'_1|$ is the scalar curvature, such that $\mathbf{e}_2(u)$ is a unit-vector pointing in the direction of the center-line $\mathbf{r}(t, u)$ curves towards at time t . Eq. 5.136 fully specifies the frame, as $\mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2$. In this context \mathbf{e}_1 , \mathbf{e}_2 and \mathbf{e}_3 are called the tangent, normal and binormal vectors respectively. co

Now recall the equations of motion of \mathbf{r} and \mathbf{e}_i along u , given by Eq. 5.40, which we write as

$$\mathbf{r}' = \mathbf{e}_i \theta_i \quad (5.137a)$$

$$\mathbf{e}'_j = \mathbf{e}_i \hat{\pi}_{ij}. \quad (5.137b)$$

Since $\mathbf{r}' = h\mathbf{e}_1$ we have that

$$\vec{\theta} = (h \ 0 \ 0)^T, \quad (5.138)$$

and as $\mathbf{e}'_1 = \kappa \mathbf{e}_2$ we have that $\hat{\pi}_{21} = -\hat{\pi}_{12} = \kappa$ and $\hat{\pi}_{31} = -\hat{\pi}_{13} = 0$. This leaves only one more possible parameter $\hat{\pi}_{23} = -\hat{\pi}_{32}$, to which we assign the scalar parameter τ , known as the *torsion*. We thus have

$$\hat{\pi} = \begin{pmatrix} 0 & -\kappa & 0 \\ \kappa & 0 & -\tau \\ 0 & \tau & 0. \end{pmatrix} \quad (5.139)$$

or, in vector form, $\vec{\pi} = (\tau \ 0 \ \kappa)^T$. Finally, we get the *Frenet-Serret equations*

$$\begin{aligned} \mathbf{e}'_1 &= \kappa \mathbf{e}_2 \\ \mathbf{e}'_2 &= -\kappa \mathbf{e}_1 + \tau \mathbf{e}_3 \\ \mathbf{e}'_3 &= \tau \mathbf{e}_2 \end{aligned} \quad (5.140)$$

which together with Eq. 5.135 reconstructs the framing $\tilde{\Phi} \cong (\mathbf{r}, E)$, from which get the filament configuration $\Phi \cong \mathbf{r}$.

We see that by choosing a gauge, we have effectively constrained the spatial configuration X to only take values in a subset $V \subset \mathfrak{se}(3)$, where $X = X(h, \kappa, \tau)$ is now parametrised by the three scalars (h, κ, τ) that are functions of space and time. Note that V is not a Lie subalgebra. This constraint establishes an injective mapping between V -valued fields $X : W \rightarrow V$ and \mathbb{E}^3 -valued fields $\Phi : W \rightarrow \mathbb{E}^3$. In this context (h, κ, τ) are known as the *differential invariants* of the filament, as they uniquely define the filament up to rigid body transformations.

We should stress that in the mathematical literature, the purview of the theory of moving frames are in general space curves (rather than space-time curves). Furthermore, in that context the metric h along the center-line is of little importance (and is usually not referred to as a differential invariant), and thus the arc-length parameterisation is preferred. Eq. 5.140 is therefore more commonly written in terms of arc-length derivatives

$$\begin{aligned} \partial_s \mathbf{e}_1 &= \tilde{\kappa} \mathbf{e}_2 \\ \partial_s \mathbf{e}_2 &= -\tilde{\kappa} \mathbf{e}_1 + \tilde{\tau} \mathbf{e}_3 \\ \partial_s \mathbf{e}_3 &= \tilde{\tau} \mathbf{e}_2 \end{aligned} \quad (5.141)$$

where $\tilde{\kappa} = h\kappa$ and $\tilde{\tau} = h\tau$ are the scalar curvature and torsion in the arc-length parameterisation.

Intuitively, we can visualise the motion of (\mathbf{r}, E) as a function of u , for fixed t , in the Frenet-Serret gauge as follows. We can imagine an airplane which is travelling in the direction of \mathbf{e}_1 and with the floor of the plane aligned with \mathbf{e}_3 . Its trajectory traces out a center-line $\mathbf{r}(u)$, at ‘speed’ h . The Frenet-Serret frame corresponds to a plane that navigates *pitching* (at a rate κ) and *rolling* (at a rate τ), but never *yaws*.

We now consider the kinematics along time t . The same kinematic equations of

motion Eq. 5.48 for the Cosserat rod still apply. However, substituting $\vec{\theta} = (h \ 0 \ 0)^T$ and $\vec{\pi} = (\tau \ 0 \ \kappa)^T$ into Eq. 5.48 we find the following constraints

$$\Omega_1 = (\Omega_3\tau - \Omega'_2)/\kappa \quad (5.142a)$$

$$\Omega_2 = -h^{-1}(D_u \vec{V})_3 \quad (5.142b)$$

$$\Omega_3 = h^{-1}(D_u \vec{V})_2. \quad (5.142c)$$

It is intuitive that the angular velocity of the filament, with a cross-section constrained to be aligned with the center-line, is completely determined as a function of the differential invariants and the translational velocity \vec{V} . Eq. 5.142, together with Eq. 5.48, completely specify the kinematics of the filament. Just as we have three scalar parameters (h, κ, τ) that determine the spatial configuration of the filament, the temporal evolution is determined by the three scalars in $\vec{V} = (V_1 \ V_2 \ V_3)$. Using Eq. 5.48 and Eq. 5.142, we can write the kinematic equations of motion for the filament compactly as

$$\dot{h} = V'_1 - \kappa V_2 \quad (5.143a)$$

$$\dot{\kappa} = \Omega'_3 + \tau \Omega_2 \quad (5.143b)$$

$$\dot{\tau} = \Omega'_1 - \kappa \Omega_2 \quad (5.143c)$$

$$(5.143d)$$

where $\vec{\Omega}$ is given by Eq. 5.142. The temporal evolution of the filament is thus specified by the velocity field \vec{V} alone. The derivation of the kinematic equations of motion of a filament Eq. 5.143 from the integrability of ξ have been previously found in [?].

The twisting filament

We can model a filament that can twist around its center-line by keeping the $SO(2)$ gauge freedom of $(\mathbf{e}_2, \mathbf{e}_3)$ and promoting it to a kinematic degree of freedom. The resulting degrees of freedom are then $\vec{\theta} = (h \ 0 \ 0)^T$ as before and $\vec{\pi}$ has all three of its components. The resulting constraints on $\vec{\Omega}$ are then

$$\Omega_2 = -h^{-1}(D_u \vec{V})_3 \quad (5.144a)$$

$$\Omega_3 = h^{-1}(D_u \vec{V})_2. \quad (5.144b)$$

instead of Eq. 5.142. The resulting kinematic equations of motion are thus

$$\dot{h} = V'_1 - \kappa V_2 \quad (5.145a)$$

$$\dot{\vec{\pi}} = D_u \vec{\Omega} \quad (5.145b)$$

The temporal evolution of the twisting filament is thus specified by the rate of twist Ω_1 and the velocity field \vec{V} .

In our airplane analogy, the twisting filament travels (again, as a function of u for fixed t) by rolling (at a rate π_1), yawing (at a rate π_2) and pitches (at a rate π_3).

5.4.2 Kinematically constrained dynamics

The formulation of the dynamics for kinematically constrained rods carries through less programmatically than in the case of the Cosserat rod. In general, extra care must be taken formulating the dynamics when the configuration space is a homogeneous space G/H . Issues of under- or over-determination in the equations of motion may arise due to the fact that $\dim(G/H) < \dim(G)$ if $\dim(H) \neq 0$.

Filament dynamics

For a filament, though its configuration space is $SE(3)/SO(3) \cong \mathbb{E}^3$ rather than $SE(3)$ as for the Cosserat rod, $SE(3)$ acts *transitively* on \mathbb{E}^3 . By which we mean that for any two elements $\mathbf{x}, \mathbf{y} \in \mathbb{E}^3$, there exists an element $g \in SE(3)$ such that $\mathbf{y} = g\mathbf{x}$. As before, we can define a Lagrangian density $\mathcal{L} : SE(3) \times TSE(3) \rightarrow \mathbb{R}$, with reduction $\ell : \mathfrak{se}(3) \rightarrow \mathbb{R}$, or use the Lagrange-D'Alembert principle as the starting point. The derivation of the dynamics then proceed as before in Sec. 5.2.2, and we find the same general equations of motion Eq. 5.99.

However, issues arise as the dimension of the Lie algebra $\dim(\mathfrak{se}(3)) = 6$ is larger than the configuration space $\dim(\mathbb{E}^3) = 3$. This is in principle partially ameliorated by choosing a gauge, such as the Frenet-Serret gauge, but if this is done naively then the resulting equations of motion are untenable. Consider the dynamical equations of motion that would result if we consider a filament with the Frenet-Serret frame, starting from the Lagrange-D'Alembert principle Eq. 5.98. As before, we let $S = \frac{\partial \mathcal{K}}{\partial N}$, where the reduced kinetic energy \mathcal{K} of the Cosserat rod was derived in Eq. 5.66 and we repeat it here

$$\mathcal{K}(N) = \frac{1}{2}\rho_0|\vec{V}|^2 + \frac{1}{2}\vec{\Omega}^T \mathbb{I} \vec{\Omega}. \quad (5.146)$$

Consider now, in particular, the moment balance equation for the filament

$$D_t \vec{L} = D_u \vec{M} + \vec{\theta} \times \vec{F} + \vec{m}, \quad (5.147)$$

where $\vec{L} = \mathbb{I} \vec{\Omega}$. Due to Eq. 5.142, $\vec{\Omega}$ is not an independent degree of freedom, but a function of \vec{V} and the differential invariants. Therefore Eq. 5.147 over-determines the system, and is in general an intractable constraint.

To address this issue, we first note that it stems from the particular method by which

we have parametrised the filament. In the fixed frame, the dynamics of the filament can be unambiguously written as

$$\rho_0 \ddot{\mathbf{r}} = \mathbf{F}' + \mathbf{f} \quad (5.148)$$

where we used $\mathbf{P} = \rho_0 \dot{\mathbf{r}}$, which can be seen as the 1-dimensional analogue of the Cauchy momentum equation in the Lagrangian point-of-view, where \mathbf{F} would correspond to the *first Piola-Kirchhoff stress tensor* in continuum dynamics.

In the geometric formulation presented here, we parametrise the filament in terms of its intrinsic geometry κ and τ , which corresponds to a given framing, rather than its center-line $\mathbf{r}(t, u)$ as in Eq. 5.148. This means that we must consider how the forces on the filament corresponds to moments on the moving frame $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$. However, as the filament is ultimately described as a center-line, the model does not accommodate the angular momentum of its frame as an independent degree of freedom. In the literature the issue of Eq. 5.147 is avoided by taking the overdamped limit modelling a filament immersed in a highly viscous fluid [177, 178, 227, 228, 229, 230, 231, 232], or by considering filament statics [197]. Here, we avoid this limit and present a derivation of inertial filament dynamics.

Let $\mathcal{K}(N) = \frac{1}{2}\rho_0|\vec{V}|^2$, so that we get $S = \{\vec{P}; \vec{0}\}$ where $\vec{P} = \rho_0 \vec{V}$, therefore implicitly assuming the angular momentum of the filament to be negligible. The moment balance equation then becomes

$$0 = D_u \vec{M} + \vec{\theta} \times \vec{F} + \vec{m}. \quad (5.149)$$

To find self-consistent constitutive dynamics, we consider moment balance in the absence of external moments $\vec{m} = 0$, to find

$$0 = (D_u \vec{M})_1 \quad (5.150a)$$

$$F_2 = h^{-1}(D_u \vec{M})_3 \quad (5.150b)$$

$$F_3 = -h^{-1}(D_u \vec{M})_2. \quad (5.150c)$$

Thus we see that combined we can only independently specify three of the six force and moment components in total. The kinodynamical equations of motion for the filament are thus

$$\dot{h} = V'_1 - \kappa V_2 \quad (5.151a)$$

$$\dot{\kappa} = \Omega'_3 + \tau \Omega_2 \quad (5.151b)$$

$$\dot{\tau} = \Omega'_1 - \kappa \Omega_2 \quad (5.151c)$$

$$D_t \vec{P} = D_u \vec{F} + \vec{f}, \quad (5.151d)$$

where the material derivatives are computed as before, using $\vec{\pi} = (\tau \ 0 \ \kappa)$ and $\vec{\Omega}$ given by

Eq. 5.142, and where the force must obey the conditions Eq. 5.150. We note that \vec{M} does not appear explicitly in the dynamics, and we can indeed in theory write down the three arbitrary constitutive force components \vec{F} , which would induce a moment via Eq. 5.150 that has no impact on the dynamics. This is however not how the constitutive forces and moments are defined in practice, as we will now show.

We now move on to treat conservative constitutive dynamics. As discussed in [207], when the Lie algebra, that acts transitively on the configuration space, is of a higher dimension than the configuration space the resulting Lagrangian dynamics (i.e. conservative dynamics) is under-determined. For a Cosserat rod we have $Q = \frac{\partial \mathcal{U}}{\partial X} = \{\vec{F}; \vec{M}\}^*$ which results in Eq. 5.72. In contrast, as $\vec{\theta} = (h \ 0 \ 0)^T$ and $\vec{\pi} = (\tau \ 0 \ \kappa)^T$ for a filament, the only components that are determined for a general constitutive \mathcal{U} are

$$F_1 = \frac{\partial \mathcal{U}}{\partial h} \quad (5.152a)$$

$$M_1 = \frac{\partial \mathcal{U}}{\partial \tau} \quad (5.152b)$$

$$M_3 = \frac{\partial \mathcal{U}}{\partial \kappa} \quad (5.152c)$$

leaving M_2 , F_2 and F_3 undetermined by \mathcal{U} . However we see that by having chosen a gauge, these force and moment components are given by moment balance Eq. 5.150. The conservative constitutive forces and moments on a filament are thus determined by Eq. 5.152 and Eq. 5.150 together. Note that it was necessary to let M_1 , F_2 and F_3 be undetermined by \mathcal{U} , rather than setting $M_1 = F_2 = F_3 = 0$, which would conflict with moment balance. The above is consistent with the force and moment balance equations found for the filament in the overdamped setting [227] and in statics [197].

It is standard for filament dynamics to associate an energetic cost to the square of the curvature and extension $\mathcal{U} = \frac{1}{2}\epsilon\kappa^2$. From Eq. 5.152 and Eq. 5.150 we then find, expressed in a non-moving frame

$$\mathbf{M} = \epsilon\kappa\mathbf{e}_3 \quad (5.153)$$

where ϵ is the bending stiffness. Eq. 5.153 agree with classical Bernoulli-Euler beam theory [227, 228, 231]. Note that as the notion of torsion τ appears from the gauge choice, rather than representing a physical property of the filament, a τ^2 term is usually not included in \mathcal{U} . Therefore $M_1 = M_2 = 0$ for a filament in the Frenet-Serret frame in practice.

In the overdamped limit, the results of Sec. 5.3.2 and Eq. 5.122 applies as before, under the constraints Eq. 5.138 and Eq. 5.139, Eq. 5.142 and Eq. 5.150. Setting $\vec{f} = -\gamma_T \vec{V}$ we

get

$$\dot{h} = V'_1 - \kappa V_2 \quad (5.154a)$$

$$\dot{\kappa} = \Omega'_3 + \tau \Omega_2 \quad (5.154b)$$

$$\dot{\tau} = \Omega'_1 - \kappa \Omega_2 \quad (5.154c)$$

$$\vec{V} = \mu_T D_u \vec{F}, \quad (5.154d)$$

where $\mu_T = \gamma_T^{-1}$.

Twisting filament dynamics

For a twisting filament, the angular velocity component Ω_1 is now a dynamic degree of freedom, and we therefore permit angular momentum along the tangential direction. We assume the principal axes of the moment of inertia \mathbb{I} coincide with the moving frame \mathbf{e}_i , $i = 1, 2, 3$, and that the angular momentum is negligible along the normal and binormal directions, setting $\vec{L} = (\mathbb{I}^{(1)} \Omega_1 \ 0 \ 0)^T$, where $\mathbb{I}^{(1)}$ is the eigenvalue of \mathbb{I} along \mathbf{e}_1 . The resulting kinodynamical equations of motion are

$$\dot{h} = V'_1 - \kappa V_2 \quad (5.155a)$$

$$\dot{\pi} = D_u \vec{\Omega} \quad (5.155b)$$

$$D_t \vec{P} = D_u \vec{F} + \vec{f} \quad (5.155c)$$

$$\mathbb{I}^{(1)} \dot{\Omega}_1 = (D_u \vec{M})_1 + m_1 \quad (5.155d)$$

where the constraints on the force are

$$F_2 = h^{-1} (D_u \vec{M})_3 \quad (5.156a)$$

$$F_3 = -h^{-1} (D_u \vec{M})_2, \quad (5.156b)$$

which should be compared with Eq. 5.151 and Eq. 5.150. For conservative constitutive we now have the relations

$$F_1 = \frac{\partial \mathcal{U}}{\partial h} \quad (5.157a)$$

$$M_i = \frac{\partial \mathcal{U}}{\partial \pi_i}. \quad (5.157b)$$

5.5 Conclusion

We studied the kinodynamics of a Cosserat rod subject to constitutive and body forces and moments. Kinematically, we identified its configuration space as the special Euclidean

group $SE(3)$ of translations and rotations on three-dimensional Euclidean space \mathbb{E}^3 . Using the Euler-Poincaré theorem we then found conservative force and moment balance equations. Furthermore, we constructed a generalised Lagrange-D'Alembert principle from which arbitrary non-conservative dynamics can be derived. The resulting kinodynamic equations of motion were formulated in terms of a spatial reconstruction field $X = \{\vec{\theta}; \vec{\pi}\}$ and generalised momentum field S , which take values in the Lie algebra $\mathfrak{se}(3)$ and its dual $\mathfrak{se}(3)^*$ respectively. The configuration of the Cosserat rod is thus encoded in terms of Lie algebraic quantities, as opposed to a more typical approach of expressing it in terms of its center-line \mathbf{r} and material frame E directly, where the latter correspond to Lie group-valued quantities. We showed that the Lie algebraic formulation naturally leads to kinodynamics expressed in terms of the intrinsic geometry of the rod. Derivatives of \mathbf{r} and E must be taken to find the shear and spatial rate-of-change of the material frame, but are found directly in components of X itself. Furthermore, the Lie algebraic formulation has a significant advantage as regards numerics: Equations of motion in terms of $SE(3)$ -quantities will in general lead to errors that push the system outside of the Lie group. For example, numerical errors will in general cause the material frame E to longer be orthogonal. On the other hand, as $\mathfrak{se}(3)$ is a vector space, errors accrued in the simulation of X and S can not push them out of the Lie algebra. Therefore, regardless of numerical errors, X will always exponentiate to a $SE(3)$ -valued Cosserat rod configuration.

Continuing, we proceeded to study the filament model, which we modelled as a Cosserat rod subject to kinematic constraints. We systematically derived its kinodynamic equations of motion in both the underdamped and overdamped limits. The filament model, which has a homogeneous configuration space $\mathbb{E}^3 \cong SE(3)/SO(3)$, is a paradigmatic example of how non Lie group-configured systems can be obtained by imposing kinematic constraints. In Ch. 6 we develop this further for general systems. We concluded the chapter by considering specific Cosserat rod models. In Sec. 5.3.1 we derived the constitutive stress energy of a three-dimensional but slender rod, and through a coarse-graining procedure we approximated its kinodynamics as that of a Cosserat rod. In Sec. 5.3.2 we formulated the dynamics of a Cosserat rod subject to internal and external frictional forces, and derived the corresponding equations of motion in an overdamped limit. Chapter 6 will build and develop many of the concepts introduced in this chapter, which will culminate in a framework to study the kinodynamics of general continuum systems with Lie group- or homogeneous configuration spaces.

Chapter 6

Geometric continuum mechanics on Lie group- and homogeneous configuration spaces

In Ch. 5, we derived the kinodynamical equations of motion of the Cosserat rod and filament, systems in one material dimension with configuration space $SE(3)$ and $SE(3)/SO(3)$ respectively. Here we extend this theory to systems with general material base spaces and Lie group- and homogeneous configuration spaces, which we can call *generalised Cosserat systems*. In this chapter we will study these systems and formulate their kinodynamics, using the geometric approach of the previous chapter. We call the result of this work a *generalised geometric Cosserat theory* (GGCT). Mathematically, the GGCT can be seen as a field theory of Lie group- or homogeneous space-valued fields defined over topological manifolds. The derivations of the results are provided in Sec. 6.2.1 and Sec. 6.2.2. To the best of our knowledge, these results are novel.

We have separated the exposition of the theory and its derivation, as the latter require a level of mathematical abstraction not necessary in the presentation of the GGCT itself. The primary mathematical challenge stems from allowing arbitrary topologies in the material base space, which entails that it longer admits a global set of coordinates. An example of such a system would be a closed surface, like the membrane of a cell. This necessitated further geometrical abstraction that was not required in the case of the Cosserat rod.

In Sec. 6.3 we use the GGCT to derive the equations of motion for a set of example systems. This section is called ‘Applications’, meant to be understood in the sense that our general geometric framework is *applied* to derive the kinodynamic theories of the given example systems. When these systems are well-established in the literature, we show that our framework recovers the correct equations of motion. Finally, in Sec. 6.1.4 we conclude with a summary and discussion of the GGCT.

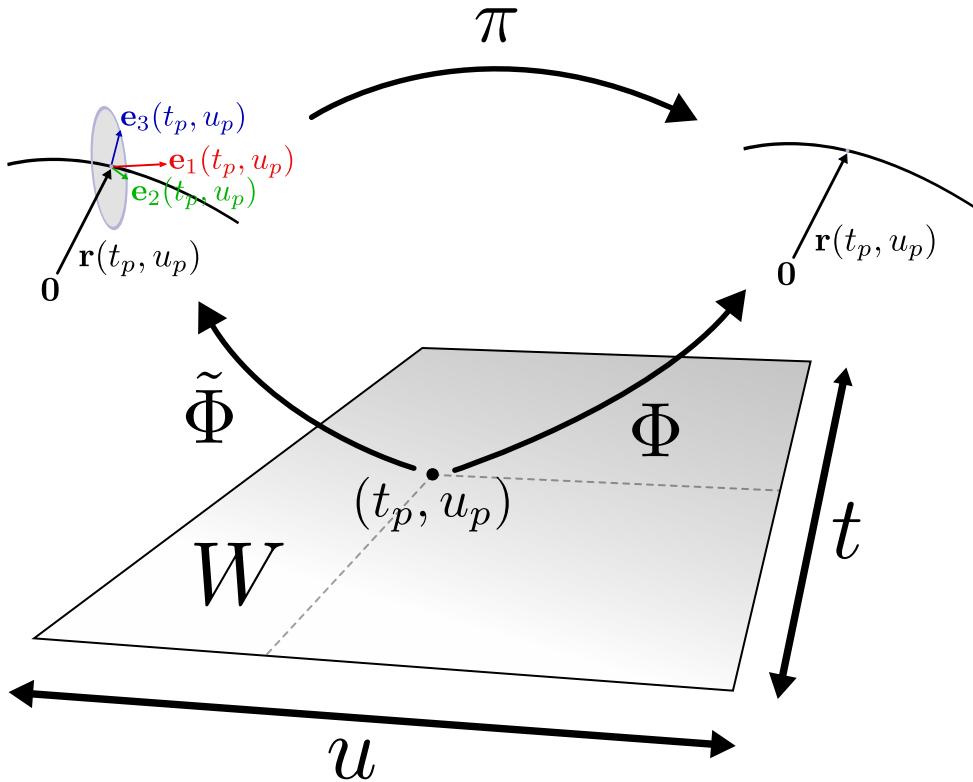


Figure 6.1: The relation between the spatio-temporal configuration Φ and the adapted frame field $\tilde{\Phi}$ of a filament, where the latter can be seen as the spatio-temporal configuration of a Cosserat rod. In the top-left and top-right we show the mapping of (t_p, u_p) , a point in the kinematic base space W , to $\tilde{\Phi}(t_p, u_p)$ and $\Phi(t_p, u_p)$, which are related by the projection map π .

6.1 Generalised geometric Cosserat theory

Here we generalise the results of Sec. 5.2 to systems with an arbitrary material base space and Lie group- or homogeneous configuration space, which we call a generalised Cosserat system. To aid understanding it will be helpful to reintroduce some nomenclature, and to relate and contextualise the mathematical formulation of the Cosserat rod and the filament to the general systems that we describe in what follows. This preamble will thus serve as a brief overview of the results given in the subsequent subsections.

The *material base space* of an open Cosserat rod is the manifold $M = [0, L_0]$, and its *kinematic configuration space* is the Lie group $SE(3)$. The *spatio-temporal configuration* of the Cosserat rod is a map $\Phi : W \rightarrow SE(3)$, where $W = [0, T] \times M$ is the *kinematic base space* and $[0, T]$ the time domain. In Sec. 5.4 we showed that a filament can be considered a kinematically constrained Cosserat rod, and its configuration space is the homogeneous space $SE(3)/SO(3)$. The spatio-temporal configuration of the filament is thus $\Phi : W \rightarrow SE(3)/SO(3)$. However, the kinematics on the homogeneous space $SE(3)/SO(3)$ can be related to kinematics on the Lie group $SE(3)$ using an *adapted frame field* (or *framing*) $\tilde{\Phi} : W \rightarrow SE(3)$ via the composition $\Phi = \pi \circ \tilde{\Phi}$, where $\pi : SE(3) \rightarrow SO(3)$

is the projection map and is defined as $\pi((\mathbf{r}; E)) = \mathbf{r}$. A given Cosserat rod configuration was thus seen as an adapted frame of the filament. This relationship is illustrated in Fig. 6.1.

The mathematical formulation of the $SE(3)$ -configured Cosserat rod, and its relation to the $SE(3)/SO(3)$ -configured filament, is analogous to the more general setting we will develop here. We first consider kinematic configuration spaces that is a general Lie group G , and we will denote \mathfrak{g} as its Lie algebra. Secondly, we let the material base space M be a general manifold. Thirdly, given a Lie subgroup $H \subset G$ known as the *stabiliser*, we can relate the kinodynamics of the G -configured system to a corresponding G/H -configured system by choosing a *gauge* in H . Therefore, as in Ch. 5, it suffices to at first consider the kinodynamics of the G -configured system with material base space M . The corresponding G/H -configured system can then be modelled by applying kinematic constraints on the G -configured system.

The spatio-temporal configurations of the general systems in consideration are therefore $\Phi : M \rightarrow G$, or $\Phi : M \rightarrow G/H$. We can thus see the generalised geometric Cosserat theory as a general theory of G - or G/H -valued fields on topological manifolds M . The majority of the mathematical technology we introduce is then for the purpose of formulating the field equations of motion on the Lie algebra \mathfrak{g} , which locally acts transitively on G/H (and on G , trivially).

In Sec. 6.1.1 and Sec. 6.1.2 we present the kinodynamic equations of motion of a general G -configured system with material base space M , formulated in terms of Lie algebraic quantities. In Sec. 6.1.3 we describe the procedure by which G/H -configured systems can be modelled via kinematically constraining the G -configured system. We conclude with a discussion of the GGCT in Sec. 6.1.4.

6.1.1 Kinematics

A physical system can be kinematically described using the GGCT if it is possible to identify its kinematic configuration at a material point with an element of a Lie group G . Alternatively, the programme is also applicable if the kinematic configuration space can be identified with a homogeneous space, upon which G acts transitively. Respectively, the Cosserat rod (Sec. 5.2) and the filament (Sec. 5.4) are examples of this respectively, and further examples can be found in Sec. 6.3. We will first treat the case where the kinematic configuration space is a Lie group, and then return to homogeneous configuration spaces in Sec. 6.1.3.

We consider a system with a d -dimensional material base space M and kinematic configuration space G , where M is a general manifold and G is an n -dimensional Lie group with Lie algebra \mathfrak{g} . Let the kinematic base space W be a $(d+1)$ -dimensional manifold of the form $W = [0, T] \times M$, where $[0, T]$ is the time domain. The spatio-temporal configuration

of the system is $\Phi : W \rightarrow G$.

As the material base space M is a general manifold, it does not admit global coordinates in general. Instead, we must have an atlas \mathcal{A} that covers M . We denote its elements as $\mathcal{A} = \{(U_a, \mathbf{u}_a) \mid a \in I\}$, where $U_a \subset M$ and $\mathbf{u}_a : M \rightarrow \mathbb{R}^d$ and I is an index set. Due to the product structure of the kinematic base space W , \mathcal{A} extends trivially to an atlas $\mathcal{A}_W = \{(U_a \times [0, T], \mathbf{x}_a) \mid a \in I\}$ over W , where $\mathbf{x}_a = (u_a^1, \dots, u_a^d, t)$. We stress that the only kinematically relevant property of M is its topological character, as the kinematics do not depend on any metric structure on M . Therefore, in practice, suitable choices for M are often d -dimensional cuboids, spheres or toruses. Conceptually, we can see M as a continuous multi-dimensional ‘index’ over the system.

The G -valued spatio-temporal configuration can be related to a \mathfrak{g} -valued 1-form as

$$\xi = \Phi^{-1} d\Phi \quad (6.1)$$

where $\xi = \tilde{\Phi}^* \omega : W \rightarrow \mathfrak{g}$ is the pull-back of the Maurer-Cartan ω form onto W , and will henceforth for the sake of brevity be referred to as the Maurer-Cartan form. In a local chart $(U, \mathbf{u}) \in \mathcal{A}$ we can write

$$\xi = N dt + X_\alpha du^\alpha \quad (6.2)$$

where $X_\alpha : [0, T] \times U \rightarrow \mathfrak{g}$, $\alpha = 1, \dots, d$ are the *spatial reconstruction fields* and $N : [0, T] \times U \rightarrow \mathfrak{g}$ is a *generalised velocity field*. Locally, we will write functions on W in terms of their coordinates, i.e. $X_\alpha(t, \mathbf{u})$.

It is important to note that the local expression of all 1-forms must transform correctly under change of charts. If ξ is written in a chart $(U', \mathbf{u}') \in \mathcal{A}$ as $\xi = N dt + X'_\alpha du'^\alpha$, then we must have

$$X'_\beta = X_\alpha \frac{\partial u^\alpha}{\partial u'^\beta} \quad (6.3)$$

on $[0, T] \times (U \cap U')$. The \mathfrak{g} -valued X_α thus transform as scalar 1-forms. If M does not admit global coordinates Eq. 6.5 must be integrated consistently over a set of charts that cover M .

The Maurer-Cartan form obeys the integrability condition

$$d\xi + \xi \wedge \xi = 0 \quad (6.4)$$

from which we can find the kinematic equations of motion expressed in a local chart

$$\partial_t X_\alpha = \mathcal{D}_\alpha N \quad (6.5)$$

where $\alpha = 1, \dots, d$, and the spatial integrability conditions

$$\partial_\beta X_\alpha = \mathcal{D}_\alpha X_\beta \quad (6.6)$$

where $\alpha = 1, \dots, d - 1$ and $\beta = \alpha + 1, \dots, d$, and where we have defined

$$\mathcal{D}_\alpha = \partial_\alpha + \text{ad}_{X_\alpha}. \quad (6.7)$$

which we call the generalised material derivative along the α th material direction, and where $\partial_\alpha = \frac{\partial}{\partial u^\alpha}$. The d spatial reconstruction fields X_α are not independent, as the spatial integrability condition Eq. 6.6 must be obeyed at all times t . However, as spatial integrability is preserved by the equations of motions Eq. 6.5, it suffices that the initial conditions satisfy Eq. 6.6 at $t = 0$. Spatial integrability also implies that the total number of degrees of freedom is $\dim(G) = d$.

To reconstruct the spatio-temporal configuration Φ at a time t from the Maurer-Cartan form ξ , we integrate the spatial part of Eq. 6.1

$$\partial_\alpha \Phi = \Phi X_\alpha, \quad \alpha = 1, \dots, d, \quad (6.8)$$

which can be integrated using a matrix ODE solver. Specifically, for a given material point \mathbf{u} and initial condition $\Phi(\bar{t}, \mathbf{u}_0)$, to find $\Phi(\bar{t}, \mathbf{u})$ we must integrate Eq. 6.8 at fixed \bar{t} along some curve $\gamma : [0, 1] \rightarrow M$, such that $\gamma(0) = \mathbf{u}_0$ and $\gamma(1) = \mathbf{u}$. See Appendix C for details on how to construct $\Phi(\mathbf{u}_0, t)$.

Due to the spatial integrability conditions Eq. 6.6, the result of integrating Eq. 6.8 is independent of the shape γ , for fixed end-points. For details on reconstructing Φ from X_α and N , see Appendix C. If the spatial reconstruction fields X_α do not satisfy spatial integrability, then solutions of Eq. 6.8 are no longer path-independent. Effectively, this means that the map $(t, \mathbf{u}) \mapsto \Phi(t, \mathbf{u})$ is not single-valued. Instead Φ is a functional $\Phi[\gamma_{\mathbf{u}}]$ of curves $\gamma_{\mathbf{u}} : [0, 1] \rightarrow M$ satisfying $\gamma_{\mathbf{u}}(0) = \mathbf{u}_0$ and $\gamma_{\mathbf{u}}(1) = \mathbf{u}$, along which X_α are integrated,¹ where \mathbf{u}_0 is some material point at which initial conditions of the exponentiation is specified.

Under the assumption that the generalised velocity field is a smooth function of \mathbf{u} , the kinematic equations of motion preserve the spatial integrability condition. However, numerical errors in the spatial integrability will in general accrue in simulations. Let

$$\Delta_{\alpha\beta}^{\text{int}}(t, \mathbf{u}) = \partial_\beta X_\alpha - \mathcal{D}_\alpha X_\beta \quad (6.9)$$

be the residual error in spatial integrability at time t and material point \mathbf{u} . In Sec. 6.2.1 we show that

$$\partial_t \Delta_{\alpha\beta}^{\text{int}} = [\Delta_{\alpha\beta}^{\text{int}}, N]. \quad (6.10)$$

We thus see that the amplitude of the residual error at any material point on the system will grow exponentially in time at any material point \mathbf{u} , if $\Delta_{\text{int}}(t, \mathbf{u}) \neq 0$. Standard integration

¹See Appendix C for details

techniques like the Forward-Euler scheme will thus in practice suffer these failures of spatial integrability. In other words, in such cases material defects may spontaneously emerge due to the numerical integration scheme. In Ch. 7 we develop geometric integrators that are designed to preserve Eq. 6.6.

6.1.2 Dynamics

The starting point for modelling the dynamics of a system using the GGCT is to identify its kinetic energy density $\mathcal{K}(\Phi)$ in terms of the spatio-temporal configuration Φ , and a reduction $\mathcal{K}(N)$ in terms of the Lie algebra-valued generalise velocity. The kinetic energy is then used to define the generalised conjugate momentum. This was done systematically for the Cosserat rod in Sec. 5.2.2. Secondly, the generalised stresses and body forces on the system must be defined. As with Cosserat rod dynamics, we can consider both constitutive and non-constitutive dynamics, as well as non-conservative dynamics. As before, the conservative case can be derived as a special case of the non-conservative case. We will be working in local material coordinates, in a chart $(U, \mathbf{u}) \in \mathcal{A}$.

The dynamic quantities correspond analogously to those in Sec. 5.2.2. Let Q^α , $\alpha = 1, \dots, d$ be the generalised constitutive stress fields, and T the generalised body force on the system, which both take values in \mathfrak{g}^* and are defined over M and can be time-dependent in general. The generalised conjugate momentum field is given by

$$S = \frac{\partial \mathcal{K}}{\partial N} \quad (6.11)$$

where $\mathcal{K} = \mathcal{K}(N)$ is a kinetic energy density defined on the system, and S takes values in the dual Lie algebra \mathfrak{g}^* , which in its matrix representation can be seen as the space $\mathfrak{g}^* = \{Y^T : Y \in \mathfrak{g}\}$. The matrix derivative is carried out using numerator-layout convention Eq. 4.70. The general dynamic equations of motion of the system are then

$$\mathcal{D}_t^* S = \mathcal{D}_\alpha^* Q^\alpha + T \quad (6.12a)$$

$$n_\alpha Q^\alpha = 0, \quad \text{on } \partial M. \quad (6.12b)$$

where $n \in T^*M$ is a normal covector field that is tangent to the material boundary ∂M and where

$$\mathcal{D}_t^* = \partial_t + \text{ad}_N^* \quad (6.13a)$$

$$\mathcal{D}_\alpha^* = \partial_\alpha + \text{ad}_{X^\alpha}^*, \quad (6.13b)$$

which we call the dual generalised material derivatives along time and the α th material directions respectively. To give an intuitive example that illustrates the implementation of

Eq. 6.12b, we can consider the case where $M = [0, L_0^1] \times [0, L_0^2] \times \cdots \times [0, L_0^d]$ and $L_0^\alpha \in \mathbb{R}^+$. Then from Eq. 6.12b we would have that $Q^\alpha = 0$ on $u^\alpha = 0, L_0^\alpha$. In physical applications S is always related linearly to N , and the kinodynamic equations of motion thus close by solving for $N = N(S)$. The total number of degrees of freedom of the system is thus $2d$, which are encoded in the spatial reconstruction fields X_α and N .

Note that in general the generalised body force T is a function of the spatio-temporal configuration Φ . In this case we must reconstruct Φ from the spatial reconstruction fields X_α to compute T at any time t . See Appendix C for a detailed algorithm.

For many physical systems we often care about formulating purely constitutive and conservative dynamics. In this case the stress can be derived from a potential energy. As we did in Sec. 5.3.1 for the Cosserat rod, the starting point would be to formulate a Lagrangian density $\mathcal{L}(\Phi, d\Phi)$. If the dynamics is purely constitutive, then the Lagrangian can not be a function of the global state Φ . Rather, only deformations $d\Phi$ energetically. Therefore the Lagrangian is only a function of the tangent space of G , and we write it as $\mathcal{L}(d\Phi)$. In such cases, $\mathcal{L}(d\Phi)$ admits a reduced form $\ell(\xi)$, that satisfies $\ell(\xi) = \mathcal{L}(d\Phi)$ when $\xi = \Phi^{-1}d\Phi$. Often, the reduced Lagrangian can be written in the form

$$\ell(\xi) = \mathcal{K}(N) - \mathcal{U}(X_\alpha). \quad (6.14)$$

Applying the Euler-Poincaré theorem, we arrive at the same equations as Eq. 6.12 with the body force absent and where the generalised internal stresses are given by

$$Q^\alpha = -\frac{\partial \ell}{\partial X_\alpha} = \frac{\partial \mathcal{U}}{\partial X_\alpha}. \quad (6.15)$$

which takes values in the dual Lie algebra \mathfrak{g}^* .

An example of how to derive a body force is given in the latter half of Sec. 5.2.2, where we incorporate gravity into the Cosserat rod model. In general, if the body force is a function of the spatio-temporal configuration Φ , then for all times t we must solve the equations $\partial_\alpha \Phi = \Phi X_\alpha$ for Φ to evaluate the body force.

Note that in practice it is often the case that Lagrangians are formulated in coordinate-form as densities, as was the case in Sec. 5.2. However, if M does not admit global coordinates it is important that the densities are defined such that for any pair of charts $(U, \mathbf{u}), (U', \mathbf{u}') \in \mathcal{A}$ where $U \cap U' \neq \emptyset$ we have

$$\mathcal{L}'(\Phi, d\Phi) = |J| \mathcal{L}(\Phi, d\Phi). \quad (6.16)$$

and

$$\ell'(\xi) = |J| \ell(\xi). \quad (6.17)$$

on $[0, T] \times U \cap U'$, where $J = |\det [\frac{\partial \mathbf{u}}{\partial \mathbf{u}'}]|$, and $\frac{\partial \mathbf{u}}{\partial \mathbf{u}'}$ is the Jacobian matrix of the coordinate

transformation between the two charts.

6.1.3 Kinematic constraints

We will now consider implementing kinematic constraints on a Lie group-configured system to a model a system that is homogeneous space-configured. The reader will notice that this procedure is less programmatic than the Lie group-configured case. The process of implementing kinematic constraints requires some care in order to develop consistent and physical equations of motion. We will therefore describe the process of kinematic constraints in as much detail as is possible whilst keeping the same level of abstraction. We will repeatedly refer to the example of the filament, treated in Sec. 5.4.1, to illustrate the discussion.

We consider a system where the kinematic configuration space is a homogeneous space. Let $\Phi : W \rightarrow G/H$ be the spatio-temporal configuration, where $H \subset G$ is an r -dimensional Lie subgroup we call the *stabiliser*. The d -dimensional Lie group G is now called a *symmetry group* over the homogeneous space G/H , over which it acts transitively. There is a natural projection map $\pi : G \rightarrow G/H$ from G to G/H , given by

$$\pi(g) = gH \quad (6.18)$$

which describes G as a principal bundle over G/H . In Sec. 5.4.1, we had $G = SE(3)$, $H = SO(3)$ and the projection $\pi((\mathbf{r}; E)) = \mathbf{r}$.

An *adapted frame field*, or *framing*, of the spatio-temporal configuration is a map $\tilde{\Phi} : G \rightarrow G/H$ which satisfies $\Phi = \pi \circ \tilde{\Phi}$. The relation between Φ , $\tilde{\Phi}$ and π is summarised by the commutative map

$$\begin{array}{ccc} & G & \\ \tilde{\Phi} \nearrow & & \searrow \pi \\ W & \xrightarrow{\Phi} & G/H \end{array}$$

and Fig. 6.1. If the stabiliser is trivial $H = \{e\}$, where $e \in G$ is the identity element, then the kinematic configuration space is $G/H \cong G$ and $\tilde{\Phi} = \Phi$.

The projection map π allows us to describe the kinematics of the system in terms of Lie group motions, using adapted frame fields. All the mathematical technology introduced in Sec. 6.1.1 and Sec. 6.1.2 apply as before, but here in terms of the adapted frame field $\tilde{\Phi}$ and its Maurer-Cartan form $\xi = \tilde{\Phi}^{-1}\tilde{\Phi}$. The kinodynamic equations of motion are thus as before Eq. 6.5 and Eq. 6.12 respectively.

Ostensibly the spatial reconstruction fields X_α , $\alpha = 1, \dots, d$ and the generalised velocity field N (or, alternatively, the generalised conjugate momentum S) together comprise $2d$ independent degrees of freedom, once the spatial integrability conditions are factored in. However, the kinematic configuration space G/H is $(n - r)$ -dimensional, leaving r

un-physical and superfluous degrees of freedom under-determined by the equations of motion. In principle there is nothing that prevents us from modelling the G/H -configured system using a G -configured system (analogously, there is nothing that prevents us to model the $SE(3)/SO(3)$ -configured filament using the $SE(3)$ -configured Cosserat rod), although as the kinematics of the latter is not adapted to the intrinsic geometry of the former formulating the dynamics can be difficult. Alternatively, we can eliminate the r -superfluous degrees of freedom.

For a given Φ , the space of admissible frame fields $\tilde{\Phi}$ is equal to the space of smooth functions of the form $h : W \rightarrow H$. Explicitly, if $\tilde{\Phi}_1$ is a framing of Φ then

$$\tilde{\Phi}_2(p) = h(p)\tilde{\Phi}_1(p), \quad \forall p \in W \quad (6.19)$$

is also a frame. We call this a *gauge freedom* in the specification of $\tilde{\Phi}$. Only if H is trivial and $r = 0$ is there a unique choice of $\tilde{\Phi}$ for each Φ .

Choosing a gauge is equivalent to prescribing a one-to-one map between spatio-temporal configurations $\Phi : W \rightarrow G/H$ and adapted frames $\tilde{\Phi} : W \rightarrow G$. In Sec. 5.4 this was done by ‘locking’ \mathbf{e}_1 and \mathbf{e}_2 of the trihedron $(\mathbf{r}, E) : W \rightarrow G = SE(3)$ to its center-line $\mathbf{r} : W \rightarrow G/H$. Thus, we had the one-to-one map $\mathbf{r} \mapsto (\mathbf{r}, E(\mathbf{r}))$. We then saw that the choice of gauge on the Lie group-level induced a d -dimensional sub-vector space $V \subset \mathfrak{se}(3)$ on the Lie algebra, where the spatial reconstruction field X only takes values in V . Note that though, in the filament case, we had ‘locked’ the spatial configuration of the material frame to $\partial_u \mathbf{r}$, in principle it is also possible to lock it to the velocity $\partial_t \mathbf{r} = \mathbf{V}$. See Sec. 6.3.4 for an explicit example of this.

In general, the choice of gauge will result in only $n - r$ components of the Maurer-Cartan form ξ to be independent, which in turn means that the kinematic equations of motion yields r constraints. See Eq. 5.142 for an example. Care must be taken to ensure to eliminate superfluous dynamic degrees of freedom. In Sec. 5.4.2 we did this by setting $S_{ij} = 0$ for every vanishing component $N_{ij} = 0$ of the generalised velocity. In turn, this will lead to r constraints on the generalised stress from the dynamical equations of motion Eq. 6.12. These r constraints are consistent with the fact that we cannot specify n generalised forces independently for a system with only $n - r$ degrees of freedom.

Though all gauge choices are theoretically equivalent, some are more ‘natural’ than others. In the case of the filament, a 1-dimensional sub-manifold of \mathbb{E}^3 , we have the Frenet-Serret, as well as the *Bishop* [234, 236], frames. A generalisation of natural moving frames to arbitrary sub-manifolds of \mathbb{E}^d can be found in [207, 233, 240]. A further generalisation of the theory of moving frames to arbitrary sub-manifolds of Lie groups can be found in [208, 241, 242, 243].

6.1.4 Summary and discussion

The GGCT can be summarised programmatically. We can model a given generalised Cosserat system by taking the following steps.

1. Identify the material base space M of the system. Only the topology of M is of kinematic relevance. For example, if the system is a closed surface (like the membrane of a cellular organism), then an appropriate material base space is $M = S^2$.
2. Identify the kinematic configuration space. In general this will be a homogeneous space G/H . If H is trivial then the kinematic configuration space is the Lie group G . This entails constructing adapted frames $\tilde{\Phi} : W \rightarrow G/H$ such that the spatio-temporal configuration $\Phi = \pi \circ \tilde{\Phi}$ kinematically encodes the state of the system.
3. Relate the Lie group-valued adapted frame to the corresponding Maurer-Cartan form ξ , using $\xi = \tilde{\Phi}^{-1}d\tilde{\Phi}$. As was done for the Cosserat rod, we can expand ξ and $\tilde{\Phi}^{-1}d\tilde{\Phi}$ in terms of their constitutive sub-matrices in order to interpret the kinematic equations of motion.
4. Write down the kinematic equations of motion Eq. 6.5 of the system, formulated on the Lie algebra \mathfrak{g} .
5. Write down a kinetic energy density of the system in terms of $d\Phi$, and its corresponding reduction $\mathcal{K}(N)$. Compute the generalised conjugate momentum $S = \frac{\partial \mathcal{K}}{\partial N}$.
6. Define the generalised stresses Q^α and generalised body force T . The conservative part of the constitutive dynamics can be derived by defining a potential energy density $\mathcal{U}(\partial_\alpha \Phi)$, with reduction $\mathcal{U}(X_\alpha)$. The generalised stresses are then $Q^\alpha = \frac{\partial \mathcal{U}}{\partial X_\alpha}$.
7. Write down the dynamic equations of motion Eq. 6.12 of the system, formulated on the dual Lie algebra \mathfrak{g}^* . Note that if the generalised body force T is Φ -dependent, then Φ must be reconstructed from X_α to compute T (see Appendix C for a detailed algorithm).
8. If the stabiliser H satisfies $\dim(H) = r > 0$, then kinematic constraints should be applied to eliminate the r superfluous degrees of freedom.

Combined, the kinodynamic equations of the motion can be written as the system of equations

$$\partial_t X_\alpha = \mathcal{D}_\alpha N \tag{6.20a}$$

$$\mathcal{D}_t^* S = \mathcal{D}_\alpha^* Q^\alpha + T \tag{6.20b}$$

that close under $S = \frac{\partial \mathcal{K}}{\partial N}$, subject to the conditions

$$\partial_\beta X_\alpha = \mathcal{D}_\alpha X_\beta \quad (6.21a)$$

$$n_\alpha Q^\alpha = 0, \quad \text{on } \partial M \quad (6.21b)$$

where $\alpha = 1, \dots, d-1$ and $\beta = \alpha+1, \dots, d$, which are spatial integrability conditions and a boundary condition on the generalised stress. The differential operators in Eq. 6.20 and Eq. 6.21 are defined as

$$\mathcal{D}_t = \partial_t + \text{ad}_N \quad (6.22a)$$

$$\mathcal{D}_\alpha = \partial_\alpha + \text{ad}_N \quad (6.22b)$$

$$\mathcal{D}_t^* = \partial_t + \text{ad}_N^* \quad (6.22c)$$

$$\mathcal{D}_\alpha^* = \partial_\alpha + \text{ad}_N^* \quad (6.22d)$$

which are the generalised material derivatives and their dualisations. In the rest of this subsection we will discuss and offer some remarks on the GGCT.

Internal and external degrees of freedom. Although not necessary for the application of the GGCT, we can introduce some further conceptual distinctions for systems with Lie group-valued kinematic configurations. For many systems there is a natural distinction to be made between the *external* and *internal* degrees of freedom. For a Cosserat rod the center-line $\mathbf{r}(t, u)$, which take values in $SE(3)/SO(3) = \mathbb{E}^3$, and the material frame E , which take values in $SO(3)$, can be seen as external and internal degrees of freedom respectively. The combination of the two, the trihedron (\mathbf{r}, E) , takes values in the combination of the two spaces $SE(3)$. We can generalise this to arbitrary systems with a configuration space G . If G admits some Lie subgroup $H \subset G$, then we may deem it natural to designate the homogeneous space G/H as the *external configuration space* and H as the *internal configuration space*, or vice versa. We should note that in general G can contain many Lie subgroups, and the distinction between the external and internal configuration spaces for a given system would be a matter of convention.

Both the results of Ch. 5, as well as the various applications in Sec. 6.3 provide examples of the implementations of the above steps, as well as a demarcation of the internal and external configuration spaces. Our aim with the chosen examples was to show that the generality of the GGCT allows for modelling a very broad class of systems with relative ease. The wide applicability of the theory stems from the fact that the kinematic configurations of most physical systems can be said to take values in either a homogeneous space or a Lie group. In addition, allowing for a general material base space M enables us to model continuum systems of arbitrary topologies.

Dynamics on G/H vs. dynamics on \mathfrak{g} . It should be noted that in principle

the kinodynamics of a G/H -configured system could be formulated entirely in terms of the G/H -valued spatio-temporal configuration $\tilde{\Phi}$ and its derivative $d\tilde{\Phi}$, as opposed to on the Lie algebra \mathfrak{g} of its symmetry group G as we have done here. This is perhaps the most common way of modelling many systems in classical continuum and Cosserat mechanics [167, 185, 226, 227]. When simulating systems in this formulation, it requires the discretisation of the G - or G/H -valued system configuration. As G and G/H are always numerically embedded in \mathbb{R}^m , for some $m > \dim(G)$ or $m > \dim(G/H)$, numerical errors accrue in such a manner so that the configuration leaves the sub-manifold $G \subset \mathbb{R}^m$ or $G/H \subset \mathbb{R}^m$. The benefit of the GGCT stems from the fact that it exploits the trivialisation $TG \rightarrow G \times \mathfrak{g}$, which enables us to formulate the kinodynamics in terms of the linear vector space \mathfrak{g} , as opposed to the non-linear space G .

Furthermore, the GGCT thus naturally leads to a formulation of the dynamics in terms of the intrinsic geometry of the system. For constitutive dynamics, this is reflected in the fact that the Lagrangian admits a reduction. The reduction procedure ‘subtracts’ all global information from the dynamics, such that only differential deformations are energetically relevant. These deformations are precisely encompassed in the Maurer-Cartan form ξ . To illustrate the difference between a G/H -based approach and a \mathfrak{g} -based approach, recall the filament system defined in Sec. 5.4. A typical potential energy for a filament is quadratic in its scalar curvature κ and extension h [178, 231]

$$U = \int_0^{L_0} (k(h - 1)^2 + \epsilon\kappa^2) du, \quad (6.23)$$

where k and ϵ are constants. As we showed in Sec. 5.4 κ is a component of the Maurer-Cartan form ξ once kinematic restrictions have been imposed. Equation 6.23 can be rewritten in terms of the $\mathbb{E}^3 \cong SE(3)/SO(3)$ -valued center-line $\mathbf{r}(u)$. However as intrinsic geometry of a space-curve $\mathbf{r}(u)$ is translation-invariant, the resulting expression is not a function of \mathbf{r} but its derivatives $\partial_u \mathbf{r}$ and $\partial_u^2 \mathbf{r}$. We find

$$U = \int_0^{L_0} \left\{ k(|\partial_u \mathbf{r}| - 1)^2 + \epsilon |\partial_u (\partial_u \mathbf{r} / |\partial_u \mathbf{r}|)|^2 \right\} du. \quad (6.24)$$

Aside from the fact that Eq. 6.23 might be aesthetically preferred over Eq. 6.24, the evaluation of the latter will be highly sensitive to numerical errors due to the derivatives.

Time as a privileged axis. As our formulation of the GGCT is primarily geared towards continuum mechanics, our notation has singled-out time as a privileged axis. To remove this notational quirk, we could simply exclude the temporal part of the kinematic base space $W = [0, T] \times M \rightarrow W = M$, and thus consider configurations $\Phi : M \rightarrow G/H$.

The resulting system of equations that determines the system configuration are then

$$\mathcal{D}_\alpha^* Q^\alpha + T = 0 \quad (6.25a)$$

$$\partial_\beta X_\alpha = \mathcal{D}_\alpha X_\beta \quad (6.25b)$$

$$n_\alpha Q^\alpha = 0, \quad \text{on } \partial M \quad (6.25c)$$

$$(6.25d)$$

where $\alpha = 1, \dots, d-1$ and $\beta = \alpha+1, \dots, d$. Equation 6.20 and Eq. 6.25 are formally equivalent. The former can be recovered from the latter by setting $M = [0, T] \times \tilde{M}$, where \tilde{M} is then a given material base space. In Eq. 6.25 we do not explicitly privilege a time-direction. However, note that we could in general consider the material base space M to be some $(d+1)$ -dimensional space-time manifold.

Field theories and non-linear σ models. The GGCT can be viewed from a field-theoretic lens, in which case the spatio-temporal configuration Φ can be seen as a G/H -valued *field* over the base space M . The GGCT is thus a formulation of field dynamics in terms of the locally transitive action of the Lie algebra \mathfrak{g} , resulting in the Lie algebraic field equations Eq. 6.25. Notably though, as opposed to the vector-valued fields we often find in many field theories (e.g. electromagnetic field theory), Φ takes values ‘outside’ of the base space M . This is the reason why M does not necessitate a metric structure in the GGCT. In principle, however, it would be possible to imbue M with a metric, and couple Φ with dynamic vector fields defined on TM in a Lagrangian formulation of the dynamics.

As a concrete example to illustrate the field-theoretic perspective, consider a Lagrangian density of the form

$$\mathcal{L} = \frac{1}{2} \eta^{\mu\nu} g(\partial_\mu \Phi, \partial_\nu \Phi) \quad (6.26)$$

where g is a (in general Φ -dependent) metric on G/H and η is the Minkowski metric (or the Euclidean metric for flat space-times). Equation 6.26 is a *non-linear σ model* [244, 245], and in this context Φ is a field that takes values in *target manifold* G/H . If Eq. 6.26 admits a reduction, then the GGCT can be applied to derive the equations of motion for Φ . This would result in Eq. 6.25, with $Q^\alpha = -\frac{\partial \ell}{\partial X_\alpha}$, where ℓ is the reduction of \mathcal{L} . In Sec. 6.3.5 we use the GGCT to derive the field equations for the $O(3)$ non-linear σ model. Furthermore, we note that Cosserat dynamics with a Lagrangian

$$\mathcal{L} = \frac{1}{2} \vec{N}^T \mathbf{M} \vec{N} + \frac{1}{2} \vec{X}^T \mathbf{K} \vec{X}, \quad (6.27)$$

of which the constitutive dynamics described in Sec. 5.3.1 is an example, is in the form of Eq. 6.26. Lagrangian Cosserat dynamics with a quadratic potential energy is thus a non-linear σ model.

Soft modes. Consider the case when the dynamics of the system is described by a constitutive Lagrangian \mathcal{L} which admits a reduction $\ell(\xi)$. We can note that the existence of the reduction $\ell(\xi)$ implies that the Lagrangian can only be dependent on the tangent space TG . In other words, \mathcal{L} has only gradient terms, and no ‘mass’ terms. Note that the components of the spatial reconstruction fields $X_\alpha = \Phi^{-1}\partial_\alpha\Phi$ are the ‘gradients’ of Φ , and therefore potential energies $\mathcal{U}(X_\alpha)$ are thus by construction ‘massless’. Consequently, each kinematic degree of freedom in such systems behave like *soft modes* [246, 247]. This can be understood intuitively by considering the Cosserat rod and the elastic energy derived in Sec. 5.3.1. The elastic energy cost of long wavelength deformations of the rod (whether twist, extension, bend or shear) go continuously to zero, as the wavelength is taken to infinity. This gives rise to sound modes (from extension), shearing waves and curvature waves.

Spatial integrability and numerics. If the spatial reconstruction fields X_α do not satisfy the spatial integrability conditions Eq. 6.21a, then the spatio-temporal configuration Φ is not single-valued as a function of (t, \mathbf{u}) . Physically, this would correspond a material defect on the system. In principle, the kinematic equations of motion we have presented can thus be used to study defects, although a careful treatment of this topic is beyond the scope of this text. However, if the generalised forces that the system is subject to are smooth, then no defects can spontaneously emerge. Nevertheless, as discussed in Sec. 6.1.1, though the kinematic equations of motion Eq. 6.20a preserve the spatial integrability conditions Eq. 6.21a, any errors in numerical simulations will grow exponentially as Eq. 6.10. In Ch. 7 we construct geometric integrators that preserve the spatial integrability.

6.2 Derivation of the geometric kinodynamical equations of motion

Here we provide detailed derivations of the results in Sec. 6.1. For the sake of clarity, we had presented the kinodynamics of Lie group-configured and homogeneous space-configured systems separately, and had formulated the latter as a kinematically constrained version of the former. Here, we will treat both cases simultaneously by considering a general system with a homogeneous configuration space. Note that for a Lie group G and $H = \{e\}$, where $e \in G$ is the identity element, then $G/H \cong G$. Therefore G is trivially a homogeneous space.

6.2.1 Kinematics

Let the kinematic base space W be a $(d + 1)$ -dimensional manifold of the form $W = [0, T] \times M$, where M is the d -dimensional material base space and $[0, T]$ is the time

domain. The spatio-temporal configuration of the system is the map $\Phi : W \rightarrow G/H$, where G is an n -dimensional symmetry group on G/H , with Lie algebra \mathfrak{g} , and the stabiliser $H \subset G$ is a r -dimensional Lie subgroup. The kinematic configuration space is the $(n - r)$ -dimensional homogeneous space G/H , upon which G acts transitively.

Let \mathcal{A} be an atlas over the material base space M , with elements $\mathcal{A} = \{(U_a, \mathbf{u}_a) \mid a \in I\}$, where $U_a \subset M$ and $\mathbf{u}_a : M \rightarrow \mathbb{R}^d$ and I is an index set. Due to the product structure of W , \mathcal{A} extends trivially to an atlas $\mathcal{A}_W = \{([0, T] \times U_a, \mathbf{x}_a) \mid a \in I\}$ over the kinematic base space W , where $\mathbf{x}_a = (t, u_a^1, \dots, u_a^d)$.

We define a projection map $\pi : G \rightarrow G/H$ as

$$\pi(g) = gH \quad (6.28)$$

which describes G as a principal bundle over G/H . Given the projection π , we can write the spatio-temporal configuration as $\Phi = \pi \circ \tilde{\Phi}$, where we $\tilde{\Phi} : W \rightarrow G$ is then an adapted frame field of Φ . In general there is no unique choice of $\tilde{\Phi}$ for a given Φ . In the case where the stabiliser is the trivial group $H = \{e\}$, where $e \in G$ is the identity element, then $G/H \cong G$ and we must have $\Phi = \tilde{\Phi}$. This was the case for the Cosserat rod.

The kinematics will be formulated with respect to the frame field $\tilde{\Phi}$, after which the projection π can be used to construct the spatio-temporal configuration Φ . The goal of this subsection is thus to formulate a mathematical programme with which we can kinematically construct $\tilde{\Phi}$. That is, given initial conditions on the time-slice at the initial time boundary, and a velocity field defined over W , we want to compute $\tilde{\Phi}$ at all future times. The vector field $d\tilde{\Phi} : W \rightarrow TG$ contains the infinitesimal information required to reconstruct Φ . As discussed in Sec. 4.2.6, we can left-translate $d\tilde{\Phi}$ to relate it to the Lie algebra-valued vector field

$$\xi = \tilde{\Phi}^{-1} d\tilde{\Phi} \quad (6.29)$$

where $\xi = \tilde{\Phi}^* \omega : W \rightarrow \mathfrak{g}$. Compactly, we write ξ locally as

$$\xi = Z_\gamma dx^\gamma \quad (6.30)$$

where $\gamma = 0, \dots, d$, such that $Z_0 = N$ and $Z_\alpha = X_\alpha$, $\alpha = 1, \dots, d$. As shown in Sec. 4.2.6, the Maurer-Cartan form satisfies the integrability condition

$$d\xi + \xi \wedge \xi = 0. \quad (6.31)$$

Substituting Eq. 6.2 into Eq. 6.31, locally we get the equations

$$\dot{X}_\alpha = \mathcal{D}_\alpha N, \quad \alpha = 1, \dots, d \quad (6.32a)$$

$$\begin{aligned} \partial_\beta X_\alpha &= \mathcal{D}_\alpha X_\beta, & \alpha &= 1, \dots, d-1, \\ && \beta &= \alpha + 1, \dots, d, \end{aligned} \quad (6.32b)$$

where we used the linear independence of the 2-form basis $du^\alpha \wedge du^\beta$ and $du^\alpha \wedge dt$, and where

$$\mathcal{D}_\alpha = \partial_\alpha + \text{ad}_{X_\alpha} \quad (6.33)$$

is the generalised material derivative along the α th material direction. Eq. 6.32 are a set of $(d+1)d/2$ conditions on ξ , which must be simultaneously satisfied. Eq. 6.32b can be seen as spatial integrability conditions on X_α , which must be satisfied at all times t , and Eq. 6.32a are then the kinematic equations of motion. To see that Eq. 6.32b and Eq. 6.32a are compatible, we take the time-derivative of the former to get

$$\begin{aligned} &\partial_t (\partial_\beta X_\alpha - (\partial_\alpha + \text{ad}_{X_\alpha}) X_\beta) \\ &= \partial_\beta \dot{X}_\alpha - \partial_\alpha \dot{X}_\beta - [\dot{X}_\alpha, X_\beta] - [X_\alpha, \dot{X}_\beta] \\ &= (\partial_\beta + \text{ad}_{X_\beta}) \dot{X}_\alpha - (\partial_\alpha + \text{ad}_{X_\alpha}) \dot{X}_\beta \\ &= \partial_\beta ([X_\alpha, N]) + [X_\beta, \partial_\alpha N] + [X_\beta, [X_\alpha, N]] \\ &\quad - \partial_\alpha ([X_\beta, N]) - [X_\alpha, \partial_\beta N] - [X_\alpha, [X_\beta, N]] \\ &= -[\partial_\alpha X_\beta, N] + [X_\beta, [X_\alpha, N]] \\ &\quad + [\partial_\beta X_\alpha, N] - [X_\alpha, [X_\beta, N]] \\ &= [\partial_\beta X_\alpha, N] - [\partial_\alpha X_\beta, N] + [[X_\beta, X_\alpha], N] \end{aligned} \quad (6.34)$$

where we used the Jacobi identity $[A, [B, C]] = -[C, [A, B]] - [B, [C, A]]$. Finally, we get

$$\partial_t \Delta_{\alpha\beta}^{\text{int}}(t, \mathbf{u}) = [\Delta_{\alpha\beta}^{\text{int}}, N]. \quad (6.35)$$

where

$$\Delta_{\alpha\beta}^{\text{int}}(t, \mathbf{u}) = \partial_\beta X_\alpha - \mathcal{D}_\alpha X_\beta \quad (6.36)$$

is the residual error in spatial integrability at time t and material point \mathbf{u} . If X_α satisfies Eq. 6.32b at time $t = 0$, then the right-hand side of Eq. 6.35 vanishes. Therefore we see that the kinematic equations of motion Eq. 6.32a preserves the spatial integrability conditions Eq. 6.32b at all future times. Equation 6.35 shows how that the amplitude of the error in spatial integrability grows exponentially in time.

The frame field $\tilde{\Phi}$ can thus be kinematically constructed by first integrating Eq. 6.32a to find ξ , and then solving by Eq. 6.29. The spatio-temporal configuration can be found

using the projection map π , as $\Phi = \pi \circ \tilde{\Phi}$.

6.2.2 Dynamics

We now consider the dynamics of a general G/H -configured system with material base space M . The derivation requires a higher level of mathematical abstraction than that of Sec. 5.2, as we are considering material bases spaces that do not admit global coordinates in general.

We will begin by first treating the case of purely-constitutive and conservative dynamics, and then proceed to include non-conservative dynamics and body forces. We will make use of the celebrated Euler-Poincaré equation, which was first derived in [214, 217] to consider the dynamics of G/H -configured point particles (that is, M is a zero-dimensional manifold). Here we apply the Euler-Poincaré theorem to general G/H -configured continuum systems. This has previously been done in the specific case of the Cosserat rod in [210].

As was analogously the case in Sec. 5.4.2 for the Cosserat rod and filament, there is no difference in the resulting equations of motion whether kinematic configuration space is G or G/H . The dynamics is formulated with respect to the transitive action of G on G/H , and therefore the generalised forces will always ostensibly be \mathfrak{g}^* -valued. Consequently, if kinematic constraints are not explicitly implemented, the resulting dynamical equations of motion will be undetermined (as was also stated in [214, 217]).

Conservative dynamics

The conservative and constitutive dynamics of the system can be described by a volume form $L(d\Phi)$ on W we call the *Lagrangian*. In local coordinates $(U, \mathbf{u}) \in \mathcal{A}$ we write

$$L(d\Phi) = \mathcal{L}(d\Phi) dt \wedge d\mathbf{u} \quad (6.37)$$

on U , where $\mathcal{L}(d\Phi)$ is a Lagrangian density, where $d\mathbf{u} = \bigwedge_{\alpha=1}^d du^\alpha$.

Note that in contrast to Sec. 5.2, where we used Lagrangian densities, here the Lagrangian is a volume form. However, in practice the Lagrangian is often defined locally in terms of densities. These densities must follow the appropriate transformation law under change of charts. Let $(U, \mathbf{u}), (U', \mathbf{u}') \in \mathcal{A}$ where $U \cap U' \neq \emptyset$. Then we must have

$$\mathcal{L}'(d\Phi) = |J| \mathcal{L}(\Phi, d\Phi). \quad (6.38)$$

on $[0, T] \times U \cap U'$, where $J = |\det [\frac{\partial \mathbf{u}}{\partial \mathbf{u}'}]|$, and $\frac{\partial \mathbf{u}}{\partial \mathbf{u}'}$ is the Jacobian matrix of the coordinate transformation between the two charts.

The dynamical equations of motion can be found from Hamilton's principle

$$\delta \int_W L(d\Phi) = 0 \quad (6.39)$$

under variations $\Phi(t) \rightarrow \Phi(t) + \delta\Phi(t)$, where $\delta\Phi(\mathbf{u}, t) \in TG$ must vanish at the temporal boundaries. As before, we define an equivalent Hamilton's principle in terms of the reduced Lagrangian

$$\delta \int_W l(\xi) = 0 \quad (6.40)$$

under permissible variations $\delta\xi = \delta(\Phi^{-1}d\Phi)$, where $l(\xi)$ is the reduced Lagrangian and a volume form on W , and in local coordinates $(U, \mathbf{u}) \in \mathcal{A}$ is expressed as

$$l(\xi) = \ell(\xi) d\mathbf{x} \quad (6.41)$$

on $[0, T] \times U$, where $d\mathbf{x} = dt \wedge d\mathbf{u}$. We have that $\ell(\xi) = \mathcal{L}(d\Phi)$ when $\xi = \Phi^{-1}d\Phi$. Analogously to Sec. 5.2.2, we find

$$\delta\xi = d\eta + \text{ad}_\xi\eta \quad (6.42)$$

where $\eta : W \rightarrow \mathfrak{g}$ is an arbitrary Lie algebra-valued test function. In local coordinates $(U_a, \mathbf{u}_a) \in \mathcal{A}$ we have

$$\delta\xi = \delta Z_\gamma du^\gamma \quad (6.43a)$$

$$\delta X_\alpha = \partial_\alpha\eta + \text{ad}_{X_\alpha}\eta, \quad \alpha = 1, \dots, d \quad (6.43b)$$

$$\delta N = \partial_t\eta + \text{ad}_N\eta \quad (6.43c)$$

on $[0, T] \times U$.

To express Eq. 6.39 and Eq. 6.40 in terms of local coordinates, we assume there exists a subset $\{(U_a, \mathbf{u}_a) \mid a \in J\} = \mathcal{B} \subset \mathcal{A}$, where $J \subset I$ is an index set, such that

$$\bigcup_{a \in J} \bar{U}_a = M \quad (6.44)$$

and $U_a \cap U_b = \emptyset$, $a, b \in J$ for $a \neq b$, where \bar{U}_k denotes the closure of the open set U_k . In other words, \mathcal{B} is a patchwork of charts over M , known as a *partition of unity*, such that their domains do not intersect but the union of their closures completely covers M . Then Hamilton's principle can be expressed in terms of local coordinates as

$$\delta \sum_{a \in J} \int_{U_a \times [0, T]} \mathcal{L}^{(a)}(d\Phi) d\mathbf{x}_a = 0 \quad (6.45)$$

or

$$\delta \sum_{a \in J} \int_{U_a \times [0,T]} \ell^{(a)}(\xi) d\mathbf{x}_a = 0. \quad (6.46)$$

From the Euler-Poincaré theorem [42, 214, 217] we have that the two variational principles Eq. 6.39 and Eq. 6.40 are equivalent, and we can thus use the latter to derive dynamical equations of motion expressed in the Lie algebra.

Let us now consider evaluating the variation over a single chart $(U, \mathbf{u}) \in \mathcal{A}$. Let $e_i, i = 1, \dots, n$ and $E_i, i = 1, \dots, n$ be bases for the Lie algebra \mathfrak{g} and its dual space \mathfrak{g}^* respectively, so that we can write $B = B_i e_i \in \mathfrak{g}$ and $C = C_i E_i \in \mathfrak{g}^*$ in terms of their components. We then have

$$\begin{aligned} \delta \int_{[0,T] \times U} \ell(\xi) d\mathbf{x} &= \int_{[0,T] \times U} A_i^\gamma \delta(Z_\gamma)_i d\mathbf{x} \\ &= \int_{[0,T] \times U} \langle \delta Z_\gamma, A^\gamma \rangle d\mathbf{x} \end{aligned} \quad (6.47)$$

where we sum over the index i , and where we have defined

$$A^\gamma = \frac{\partial \ell}{\partial Z_\gamma} \in \mathfrak{g}^* \quad (6.48)$$

and where the inner product $\langle \cdot, \cdot \rangle : \mathfrak{g} \times \mathfrak{g}^* \rightarrow \mathbb{R}$ is given by

$$\langle B, C \rangle = B_i C_i \quad (6.49)$$

for $B \in \mathfrak{g}$ and $C \in \mathfrak{g}^*$. We will now massage Eq. 6.47 into a coordinate-free form, so that the variation can be performed without recourse to charts.

Whilst the topology of the material base space M is of kinematic importance, as it dictates whether multiple charts are needed to parametrise the entire space, its differential geometry is irrelevant to both the kinematics as well as the dynamics. Nevertheless, for the following, we will require a volume form on M . For the derivation to proceed the choice is arbitrary, but one natural choice is the reduced Lagrangian ℓ . However, to emphasise that the choice is arbitrary we will consider a general volume form, which can be written in a local chart $([0,T] \times U, \mathbf{u}) \in \mathcal{A}_W$ as $\omega = w d\mathbf{x}$.

We define the \mathfrak{g}^* -valued vector field

$$\begin{aligned} \Gamma &= w^{-1} A^\gamma \frac{\partial}{\partial x^\gamma} \\ &= w^{-1} (S^{(a)} \partial_t - Q_\alpha^{(a)} \partial_\alpha) \end{aligned} \quad (6.50)$$

such that

$$\begin{aligned} i_\Gamma \omega &= \sum_{\gamma=0}^d (-1)^\gamma A^\gamma \bigwedge_{\kappa \neq \gamma} dx^\kappa \\ &= S^{(a)} d\mathbf{u}_a + \sum_{\alpha=1}^d (-1)^\alpha Q_\alpha^{(a)} dt \wedge \left(\bigwedge_{\alpha \neq \beta} du_a^\beta \right) \end{aligned} \quad (6.51)$$

where $i_\Gamma \omega$ denotes the interior product of Γ and ω , and we have written $S = A^0$ and $Q^\alpha = A^\alpha$, $\alpha = 1, \dots, d$. It can be verified that Eq. 6.47 transforms as a vector under change of charts, using Eq. 6.38 and Eq. 6.3.

We can now rewrite Eq. 6.47 as

$$\int_{[0,T] \times U} \langle \delta\xi \wedge i_\Gamma \omega \rangle = 0 \quad (6.52)$$

where we have defined the inner wedge product

$$\langle B \wedge C \rangle = B_i \wedge C_i. \quad (6.53)$$

for \mathfrak{g} - and \mathfrak{g}^* -valued 1-forms B and C respectively. Now as $\delta\xi$ and $\langle \delta\xi \wedge i_\Gamma \omega \rangle$ are 1- and $(d+1)$ -forms respectively, $i_\Gamma \omega$ must be a \mathfrak{g}^* -valued d -form. Since Eq. 6.52 is chart-independent we can write Hamilton's principle as

$$\int_W \langle \delta\xi \wedge i_\Gamma \omega \rangle = 0. \quad (6.54)$$

Now using Eq. 4.20 we have

$$d\langle \eta, i_\Gamma \omega \rangle = \langle d\eta \wedge i_\Gamma \omega \rangle + \langle \eta, di_\Gamma \omega \rangle \quad (6.55)$$

with which we can perform integration-by-parts, to get

$$\begin{aligned} \int_W \langle \delta\xi \wedge i_\Gamma \omega \rangle &= \int_W \{ \langle d\eta \wedge i_\Gamma \omega \rangle + \langle \text{ad}_\xi \eta \wedge i_\Gamma \omega \rangle \} \\ &= \int_W \{ d\langle \eta, i_\Gamma \omega \rangle - \langle \eta, di_\Gamma \omega \rangle - \langle \eta, \text{ad}_\xi^*(i_\Gamma \omega) \rangle \} \\ &= \int_{\partial W} \langle \eta, i_\Gamma \omega \rangle - \int_W \{ \langle \eta, (d + \text{ad}_\xi^*) i_\Gamma \omega \rangle \} \\ &= 0 \end{aligned} \quad (6.56)$$

where we used Stokes' theorem Eq. 4.27, and $\text{ad}_B^* : \mathfrak{g}^* \rightarrow \mathfrak{g}^*$ is the dual of the adjoint action defined as $\langle \text{ad}_D B, C \rangle = \langle B, \text{ad}_D^* C \rangle$ for $D, B \in \mathfrak{g}$ and $C \in \mathfrak{g}^*$.

For arbitrary η that vanishes at the temporal boundaries, the integral Eq. 6.56 must

vanish. We first consider the boundary term in local coordinates

$$\begin{aligned}
& \int_{\partial W} \left\{ \langle \eta, S \rangle d\mathbf{u} + \sum_{\alpha=1}^d (-1)^\alpha \langle \eta, Q^\alpha \rangle dt \wedge \left(\bigwedge_{\alpha \neq \beta} du^\beta \right) \right\} \\
&= \int_{\partial W} \sum_{\alpha=1}^d (-1)^\alpha \langle \eta, Q^\alpha \rangle dt \wedge \left(\bigwedge_{\alpha \neq \beta} du^\beta \right) \\
&= \int_0^T dt \int_{\partial M} \sum_{\alpha=1}^d (-1)^\alpha \langle \eta, Q^\alpha \rangle \bigwedge_{\alpha \neq \beta} du^\beta
\end{aligned} \tag{6.57}$$

where we used that η vanishes at the temporal boundaries. The integral over ∂W and ∂M in Eq. 6.57 is an abuse of notation, and meant to be understood as an integral over the boundary covered by the chart. Consider the tangent and cotangent spaces $T_p M$ and $T_p^* M$ at a point $p \in \partial M$, and we trivially consider $T_p \partial M$ as a subset $T_p \partial M \subset T_p M$. Up to scalar multiplication, there is a single covector $n_p \in T_p^* M$ such that $n_p(v) = 0$ for all $v \in T_p \partial M$. We extend this to all $p \in \partial M$, to define the normal covector field n . Now let us assume that the chart $(U, \mathbf{u}) \in \mathcal{A}$ is defined such that the coordinates of the boundary satisfies $u^1 = 0$ and the interior $u^1 > 0$. In other words, we have that $(u^1)^{-1}(0) = \partial W \cap \bar{U}$. In these coordinates we have that $n_1 = -1$ and $n_\alpha = 0$, $\alpha \neq \kappa$. Eq. 6.57 can then be rewritten as

$$\begin{aligned}
& \int_0^T dt \int_{\partial M} (-1)^\kappa \langle \eta, Q^1 \rangle \bigwedge_{\alpha \neq \kappa} du^\beta \\
&= \int_0^T dt \int_{\partial M} \langle \eta, Q^1 \rangle du^2 \dots du^d \\
&= \int_0^T dt \int_{\partial M} n_\alpha \langle \eta, Q^\alpha \rangle du^2 \dots du^d
\end{aligned} \tag{6.58}$$

where the integrals in the second and third lines are the standard integral in multi-variate calculus, and we sum over α in the third line. As Eq. 6.58 must vanish for arbitrary η , we find the boundary condition

$$n_\alpha Q^\alpha = 0, \tag{6.59}$$

in the given chart. To see that Eq. 6.59 is chart-independent, note that $n_\alpha = \frac{\partial u'^\beta}{\partial u^\alpha} n'_\beta$ and that Q_α transforms as a vector density

$$Q_\alpha = \frac{\partial \ell}{\partial X_\alpha} = |J| \frac{\partial \ell'}{\partial X'_\beta} \frac{\partial X'_\beta}{\partial X_\alpha} = |J| Q'_\beta \frac{\partial u^\alpha}{\partial u'^\beta} \tag{6.60}$$

where $J = \det \left[\frac{\partial \mathbf{u}'}{\partial \mathbf{u}} \right]$ and we used Eq. 6.3 and Eq. 6.38. We thus have that $n_\alpha Q_\alpha = |J| n'_\alpha Q'_\alpha$, which preserves Eq. 6.59 as $|J| \neq 0$.

From the second term in the Eq. 6.56 we find

$$\mathcal{D}^*(i_\Gamma\omega) = 0 \quad (6.61a)$$

which are the dynamical equations of motion in invariant form, where we have defined

$$\mathcal{D}^* = d + \text{ad}_\xi^*. \quad (6.62)$$

By substituting Eq. 6.51 into Eq. 6.61 we get the equations of motion in a local chart

$$\mathcal{D}_t^* P = \mathcal{D}_\alpha^* Q^\alpha \quad (6.63a)$$

$$n_\alpha Q^\alpha = 0, \quad \text{on } \partial M. \quad (6.63b)$$

where we sum over $\alpha = 1, \dots, d$, and have defined

$$\mathcal{D}_t^* = \partial_t + \text{ad}_N^* \quad (6.64a)$$

$$\mathcal{D}_\alpha^* = \partial_\alpha + \text{ad}_{X_\alpha}^* \quad (6.64b)$$

which we call the dual generalised material derivatives.

Volumetric force forms and non-conservative dynamics

Here we generalise the above to include non-conservative constitutive dynamics and volumetric forces. In Sec. 5.2.2 we called the latter the *body forces* and *body moments* on the Cosserat rod. Here the analogous mathematical object is a *covector*-valued volume form, which we call the generalised body force density on the system.

We start with the *Lagrange-D'Alembert principle*, by adding an additional term to Eq. 6.39

$$\delta \int_W L(d\Phi) + \int_W (\mathbf{T}(\Phi, d\Phi, \dots))(\delta\Phi) = 0 \quad (6.65)$$

where \mathbf{T} is a T^*G -valued volume form we call the *generalised volumetric force*, and is acting on $\delta\Phi \in \Gamma(TG)$. Henceforth we will suppress its arguments and write $(\mathbf{T}(\Phi, d\Phi, \dots))(\delta\Phi) = \mathbf{T}(\delta\Phi)$. The second-term in Eq. 6.65 is analogous to external forces in classical Euler-Poincaré theory for point-particles [42, 217]. In particular, compare Eq. 6.65 with the integral Lagrange-D'Alembert principle for point-particles in [42, Eq. 7.8.5].

As before we want to express Eq. 6.65 in reduced form. As in Sec. 5.2.2 we have that $\mathbf{T}(\delta\Phi) = \langle \eta, T \rangle$, where $T = L_\Phi^* \mathbf{T} = (\Phi^T)^{-1} \mathbf{T}$ is the generalised volumetric force and is a \mathfrak{g}^* -valued volume form, and $L_\Phi^* : T_\Phi^* G \rightarrow \mathfrak{g}^*$ is a mapping from the cotangent bundle to the dual Lie algebra defined as $\mathbf{T}(L_\Phi \eta) = \langle L_\Phi^* \mathbf{T}, \eta \rangle$. From the reduced form of Eq. 6.65,

and following the same steps in Eq. 6.56, we get

$$\int_{\partial W} \langle \eta, i_\Gamma \omega \rangle - \int_W \{ \langle \eta \wedge \mathcal{D}^* i_\Gamma \omega \rangle \} + \int_W \langle \eta, T \rangle = 0. \quad (6.66)$$

and the equations of motion

$$\mathcal{D}^*(i_\Gamma \omega) = T \quad (6.67a)$$

with the same boundary conditions as before. In local coordinates we have

$$\mathcal{D}_t^* S = \mathcal{D}_\alpha^* Q^\alpha + T \quad (6.68a)$$

$$n_\alpha Q^\alpha = 0, \quad \text{on } \partial M. \quad (6.68b)$$

The argument for how to include non-conservative constitutive dynamics mimics that of Sec. 5.2.2. Here, we need only consider a more general Eq. 6.51, where Q^α is non-variational in general. We can thus write down a *generalised Lagrange-D'Alembert principle* for continuum systems as

$$\int_W \langle \delta \xi \wedge i_\Gamma \omega \rangle + \int_W \langle \eta, T \rangle = 0. \quad (6.69)$$

where the generalised internal stresses Q^α are non-variational in general for each $a \in I$, and η must vanish at the temporal boundaries. In terms of local charts we have

$$\int_{[0,T] \times U} \{ \langle \delta N, S \rangle - \langle \delta X_\alpha, Q^\alpha \rangle \} d\mathbf{x} + \int_{[0,T] \times U} \langle \eta, T \rangle = 0 \quad (6.70)$$

where we sum over $\alpha = 1, \dots, d$, and δN and δX_α are given in Eq. 6.43, from which general non-variational dynamics, including constitutive and body forces, can be derived.

6.3 Applications

Here we will apply the geometric framework derived in the preceding sections to derive the equations of motion of various example systems. In contrast to the careful treatment of the Cosserat rod in Sec. 5.2, we will not carry out derivations in details. Rather, we will take the results of Sec. 6.1 as a given and apply them directly. In doing so we mean to emphasise that our framework is programmatic, flexible and applicable to a wide variety of systems.

Sections 6.3.1 and 6.3.2 derive the classical theories of Cosserat bodies and surfaces, which are analogous to the Cosserat rod in two and three material dimensions respectively. Furthermore, we show that when the internal angular momenta are neglected, these models

reduce to the theory of classical continuum mechanics, but expressed in terms of the intrinsic geometry of the bodies. Sections 6.3.3 and 6.3.4 provide examples of generalised Cosserat rods, where the external configuration spaces are not \mathbb{E}^3 , but the 2-sphere and Minkowski space respectively. When appropriate, we will point out if there is a natural division between internal and external degrees of freedom in the kinematic configuration spaces, which was a distinction that was introduced in Sec. 6.1.4. In Sec. 6.3.5 we present a final example where we apply the GGCT to the $O(N)$ non-linear σ model, as a showcase and illustration for the connection between our formalism and field theory

6.3.1 Three-dimensional Cosserat media

In Sec. 5 we approximated the kinodynamics of a slender rod with the Cosserat rod model, which in that setting was seen as a coarse-graining of a three-dimensional slender rod into a one-dimensional deformable rod of rigid body cross-sections. In general however, classical Cosserat media can exist in up to three spatial dimensions, and their point-continua can have an arbitrary number of directors, as discussed in Sec. 4.1. Here we consider a Cosserat *body* in three material dimensions (u, v, w) with two orthonormal directors, where the latter will be represented as material frames attached at each material point (u, v, w) . This is thus analogous to the Cosserat rod in three material dimensions. However, here the micro-structure of the point-continua are not the result of a coarse-graining procedure, but rather represent true internal degrees of freedom. Examples of such systems include liquid crystals [189, 190] as well as electromagnetic and ferromagnetic media [191, 192, 193].

Let the material base space be a connected and bounded subset $M \subset \mathbb{R}^3$. For example $M = [0, L_0^u] \times [0, L_0^v] \times [0, L_0^w]$ would be an appropriate material base space for a system with a cuboidal rest configuration. As for the Cosserat rod, the total configuration space is $SE(3)$, and the external and internal configuration spaces are $\mathbb{E}^3 = SE(3)/SO(3)$ and $SO(3)$ respectively. Let $\mathbf{u} = (u, v, w) : M \rightarrow \mathbb{R}^3$ be global material coordinates on the body, and let $\mathbf{r} : [0, T] \times M \rightarrow \mathbb{E}^3$ be the spatial configuration of the continuum body at time t , and let $E(t, \mathbf{u}) = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \in SO(3)$ be the material frame at the material point \mathbf{u} at time t . As before, we have

$$\Phi(t, \mathbf{u}) = (\mathbf{r}(t, \mathbf{u}); E(t, \mathbf{u})) = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{r}(t, \mathbf{u}) & E(t, \mathbf{u}) \end{pmatrix} \quad (6.71)$$

and we write the Maurer-Cartan form as

$$\xi = Ndt + X_u du + X_v dv + X_w dw \quad (6.72)$$

where

$$N = \{\vec{V}; \vec{\Omega}\} \quad (6.73a)$$

$$X_u = \{\vec{\theta}_u; \vec{\pi}_u\} \quad (6.73b)$$

$$X_v = \{\vec{\theta}_v; \vec{\pi}_v\} \quad (6.73c)$$

$$X_w = \{\vec{\theta}_w; \vec{\pi}_w\}. \quad (6.73d)$$

From $\Phi^{-1}d\Phi = \xi$ we find that

$$d\mathbf{r} = \mathbf{V}dt + \boldsymbol{\theta}_u du + \boldsymbol{\theta}_v dv + \boldsymbol{\theta}_w dw \quad (6.74a)$$

$$dE = E\hat{\Omega}dt + E\hat{\pi}_u du + E\hat{\pi}_v dv + E\hat{\pi}_w dw. \quad (6.74b)$$

From Eq. 6.5 we have that the kinematic equations of motion are

$$\dot{X}_\alpha = \mathcal{D}_\alpha N \quad (6.75a)$$

where $\alpha = u, v, w$, and where the spatial reconstruction fields X_α must satisfy the $d(d+1)/2 = 6$ spatial integrability conditions Eq. 6.6 at all times. As spatial integrability is preserved by the forward equations Eq. 6.75, it suffices that X_α satisfy Eq. 6.6 at $t = 0$. Substituting Eq. 6.73 into Eq. 6.75 we get

$$D_t \vec{\theta}_\alpha = D_\alpha \vec{V} \quad (6.76a)$$

$$\dot{\vec{\pi}}_\alpha = D_\alpha \vec{\Omega} \quad (6.76b)$$

where $D_t = \partial_t + \hat{\Omega}$ and $D_\alpha = \partial_\alpha + \hat{\pi}_\alpha$, $\alpha = u, v, w$.

From Eq. 6.12 we find the dynamical equations of motion

$$\mathcal{D}_t^* S = \mathcal{D}_u^* Q^u + \mathcal{D}_v^* Q^v + \mathcal{D}_w^* Q^w + T \quad (6.77a)$$

$$n_\alpha Q^\alpha = 0, \quad \text{on } \partial M. \quad (6.77b)$$

where ad_*^* was given in Eq. 5.81, and where

$$S = \frac{\partial \mathcal{K}}{\partial N} = \{\vec{P}; \vec{L}\}^* \quad (6.78a)$$

$$Q^u = \{\vec{F}^u; \vec{M}^u\}^* \quad (6.78b)$$

$$Q^v = \{\vec{F}^v; \vec{M}^v\}^* \quad (6.78c)$$

$$Q^w = \{\vec{F}^w; \vec{M}^w\}^* \quad (6.78d)$$

$$T = \{\vec{f}; \vec{m}\}^* \quad (6.78e)$$

We now define the kinetic energy per unit material volume as

$$\mathcal{K} = \mathcal{K}(N) = \frac{1}{2}\rho_0|\vec{V}|^2 + \frac{1}{2}\vec{\Omega}^T\mathbb{I}\vec{\Omega} \quad (6.79)$$

where ρ_0 is the mass density per unit material volume and \mathbb{I} is the moment-of-inertia per unit material volume, and we have $\vec{P} = \rho_0\vec{V}$ and $\vec{L} = \mathbb{I}\vec{\Omega}$. Substituting Eq. 6.73 and Eq. 6.78 into Eq. 6.77 we get

$$D_t\vec{P} = D_\alpha\vec{F}^\alpha + \vec{f} \quad (6.80a)$$

$$D_t\vec{L} = D_\alpha\vec{M}^\alpha + \vec{\theta}_\alpha \times \vec{F}^\alpha + \vec{m} \quad (6.80b)$$

$$n_\alpha\vec{F}^\alpha = n_\alpha\vec{M}^\alpha = 0, \quad \text{on } \partial M \quad (6.80c)$$

where we sum over repeated indices of $\alpha = u, v, w$. Let L, T and M refer to the dimensions of material length, time and mass respectively. The dimensions of all kinodynamic quantities can be found by first noting that from Eq. 6.74 we have that $[\vec{V}] = LT^{-1}$, $[\vec{\theta}_\alpha] = 1$, $[\vec{\Omega}] = T^{-1}$ and $[\vec{\pi}_\alpha] = L^{-1}$. As $\mathcal{K}(N)$ has units of energy per unit material volume, and $\vec{P} = \frac{\partial\mathcal{K}}{\partial\vec{V}}$ and $\vec{L} = \frac{\partial\mathcal{K}}{\partial\vec{\Omega}}$, we have that $[\vec{P}] = ML^{-2}T^{-1}$ and $[\vec{L}] = ML^{-1}T^{-1}$. Similarly, as $\vec{F}^\alpha = \frac{\partial\mathcal{U}}{\partial\vec{\theta}_\alpha}$ and $\vec{M}_\alpha = \frac{\partial\mathcal{U}}{\partial\vec{\pi}_\alpha}$ for conservative dynamics, we have that $[\vec{F}^\alpha] = ML^{-1}T^{-2}$ and $[\vec{M}^\alpha] = MT^{-2}$. We thus conclude that \vec{P} and \vec{L} have units of momentum and angular momentum per unit material volume respectively, and \vec{F}^α and \vec{M}^α force and moment per unit material area respectively. The body force \vec{f} and \vec{m} have units of force and moment per unit material volume respectively.

Equations 6.76 and 6.80 together completely determine the kinodynamics of the Cosserat body, where the initial conditions must obey the spatial integrability constraints Eq. 6.6, and are consistent with the equations of motion found in the literature [196]. The equations close using $\vec{P} = \rho_0\vec{V}$ and $\vec{L} = \mathbb{I}\vec{\Omega}$.

We will now show that Eq. 6.80 is consistent with the dynamic laws of classical continuum mechanics. We can recognise Eq. 6.80a as the *Cauchy momentum equation* in Lagrangian coordinates with respect to a reference configuration. This is especially clear in the non-moving frame. We define $\nabla_\circ = (\partial_u \ \partial_v \ \partial_w)^T$ and

$$\Sigma = \begin{pmatrix} F^u & F^v & F^w \end{pmatrix} \quad (6.81a)$$

$$\boldsymbol{\Sigma} = E\Sigma = \begin{pmatrix} \mathbf{F}^u & \mathbf{F}^v & \mathbf{F}^w \end{pmatrix} \quad (6.81b)$$

such that $\nabla_\circ \cdot \boldsymbol{\Sigma}^T = \partial_\alpha \mathbf{F}_\alpha = ED_\alpha\vec{F}_\alpha$. Then Eq. 6.80a in the non-moving frame becomes

$$\rho_0\ddot{\mathbf{r}} = \nabla_\circ \cdot \boldsymbol{\Sigma}^T + \mathbf{f}, \quad (6.82)$$

which is the *Cauchy momentum equation*, written in terms of the *first Piola-Kirchhoff*

stress tensor Σ . Equation 6.80a can thus in turn be seen as the Cauchy momentum equation in curvilinear coordinates. The first Piola-Kirchoff stress tensor is related to the Cauchy stress tensor σ as $\mathcal{P}\Sigma^T = \det[\mathcal{P}]\sigma$ where $\mathcal{P} = \frac{\partial \mathbf{r}}{\partial \mathbf{u}}$ is the deformation gradient tensor, using which we can relate Eq. 6.82 to Eq. 4.1 Analogously, Eq. 6.80b can thus also be seen as a conservation equation for the internal angular momentum of the system.

In Eq. 6.81 we see saw that the first Piola-Kirchoff in the moving frame was Σ . Similarly, if we define

$$\Theta = \begin{pmatrix} \vec{\theta}_u & \vec{\theta}_v & \vec{\theta}_w \end{pmatrix} \quad (6.83)$$

we have that $\mathcal{P} = E\Theta$. We thus see that Θ is the deformation gradient tensor in the moving frame. Furthermore we have that if \mathcal{U} is the density of potential energy stored in the system, then since in classical continuum mechanics [248]

$$\Sigma = \frac{\partial \mathcal{U}}{\partial \mathcal{P}} \quad (6.84)$$

we have that

$$\Sigma = E^T \frac{\partial \mathcal{U}}{\partial \mathcal{P}} = \frac{\partial \mathcal{U}}{\partial \Theta} \quad (6.85)$$

which recovers $\vec{F}_\alpha = \frac{\partial \mathcal{U}}{\partial \vec{\theta}_\alpha}$.

We have thus shown that conservative Cosserat body dynamics is consistent with conservative continuum dynamics, and reduces to it when eliminating the internal degrees of freedom through kinematic restriction. This elimination is trivial in the case of the Cosserat body. As the material frame lies in the tangent space of the body at all points $\mathbf{e}_i(t, \mathbf{u}) \in T_{\mathbf{r}(t, \mathbf{u})}\mathbb{E}^3$, the frame is always adapted. Therefore one simple gauge choice is to use a non-moving frame by setting $\vec{\pi} = \vec{\Omega} = 0$, thus neglecting Eq. 6.80b

Finally, we can also relate $\vec{\theta}_\alpha$ and $\vec{\pi}_\alpha$ to well-known quantities in differential geometry. Let \tilde{d} be a ‘spatial’ exterior derivative, which we define such that $\tilde{d}\mathbf{r} = \vec{\theta}_\alpha du^\alpha$. At time t , the metric $g_{\alpha\beta}$ of the manifold $\mathbf{r}(M) \subset \mathbb{E}^3$ satisfies

$$\tilde{d}\mathbf{r} \cdot \tilde{d}\mathbf{r} = g_{\alpha\beta} du^\alpha du^\beta \quad (6.86)$$

we can thus identify $g_{\alpha\beta} = \vec{\theta}_\alpha^T \vec{\theta}_\beta$, or $g = \Theta^T \Theta$. Furthermore, in Euclidean space the *Christoffel symbols* are defined as

$$\partial_\alpha \mathbf{e}_j = \Gamma_{j\alpha}^i \mathbf{e}_i \quad (6.87)$$

we can thus identify $(\hat{\pi}_\alpha)_{ij} = \Gamma_{j\alpha}^i$, and D_α as the covariant derivative on $\mathbf{r}(M)$.

6.3.2 Cosserat surfaces

We consider a thin shell with reference configuration $\mathcal{M} = [-L_0^1/2, L_0^1/2] \times [0, L_0^2] \times [0, L_0^3]$, material coordinates $\mathbf{X} \in \mathcal{M}$ and with constant mass density per unit material volume ρ_0^V .

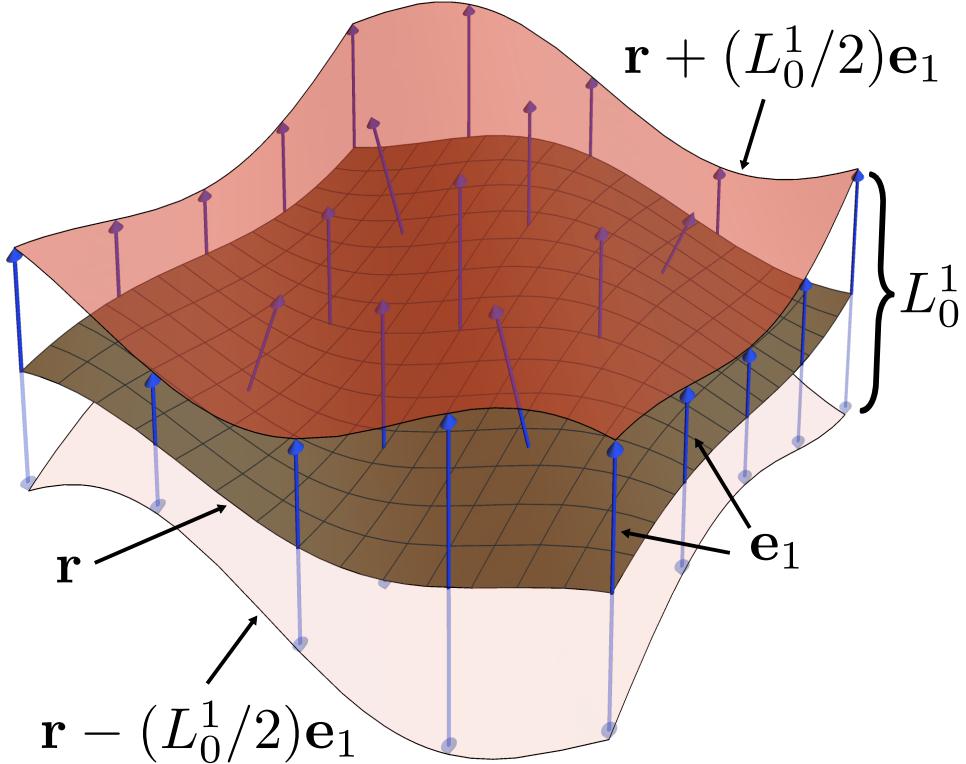


Figure 6.2: The director field $\mathbf{e}_1(u, v)$ (blue arrows) and the midsurface $\mathbf{r}(u, v)$ (brown surface) of the Cosserat surface approximates a thin shell by constraining the material fibres along the director to be fixed. The upper and lower boundaries of the shell (transparent red surfaces) are given by $\mathbf{r}(u, v) \pm (L_0^1/2)\mathbf{e}_1(u, v)$ respectively.

We assume that the body is slender along the first spatial dimension, meaning that L_0^1 is small relative to L_0^2 and L_0^3 . At time t the location of the material point at \mathbf{X} is given by $\mathbf{x}(\mathbf{X}, t)$, which are the deformed coordinates. We also define X_1 such that $\int_M X_1 d\mathbf{X} = 0$.

As we did in Sec. 5.2.2 for a slender rod, we will now kinematically approximate the thin shell with a Cosserat system of lower dimensionality. We consider a Cosserat surface, with material base space $M = [0, L_0^2] \times [0, L_0^3]$, and we denote its material coordinates as $(u, v) = (X_2, X_3)$. A single inextensible director \mathbf{e}_1 is attached at each material point (u, v) . Then, the Cosserat surface approximates the slender body under the kinematic assumption that the fibres along \mathbf{e}_1 are rigid bodies. In other words we have

$$\mathbf{x}(\mathbf{X}, t) = \mathbf{r}(t, u, v) + X_1 \mathbf{e}_1. \quad (6.88)$$

where $X_1 \in [-L_0^1/2, L_0^1/2]$ and $\mathbf{r}(t, u, v)$ is the mid-surface. Equation 6.88 is illustrated in Fig. 6.2. Notably the Cosserat surface only has a single director, as opposed to the two directors of the Cosserat rod. This means that the internal configuration space of

the Cosserat surface is $S^2 = SO(3)/SO(2)$ rather than $SO(3)$. However, for the sake of convenience we may add an additional orthonormal director \mathbf{e}_2 such that we have a full material frame $E \in SO(3)$. We thus introduce a gauge freedom in the physical description of the system, but due to the abelian nature of $SO(2)$ this will not cause any issues in the physical description of the system.

We write $\Phi(t, u, v) = (\mathbf{r}(t, u, v); E(t, u, v))$ as before and

$$\xi = Ndt + X_u du + X_v dv \quad (6.89)$$

where

$$N = \{\vec{V}; \vec{\Omega}\} \quad (6.90a)$$

$$X_u = \{\vec{\theta}_u; \vec{\pi}_u\} \quad (6.90b)$$

$$X_v = \{\vec{\theta}_v; \vec{\pi}_v\}. \quad (6.90c)$$

From $\Phi^{-1}d\Phi = \xi$ we find

$$d\mathbf{r} = \mathbf{V}dt + \boldsymbol{\theta}_u du + \boldsymbol{\theta}_v dv \quad (6.91a)$$

$$dE = E\hat{\Omega}dt + E\hat{\pi}_u du + E\hat{\pi}_v dv. \quad (6.91b)$$

and from Eq. 6.5 we find the kinematic equations of motion

$$\dot{X}_u = \mathcal{D}_u N \quad (6.92a)$$

$$\dot{X}_v = \mathcal{D}_v N \quad (6.92b)$$

where the spatial reconstruction fields X_u and X_v must obey the spatial integrability conditions

$$\partial_v X_u = \mathcal{D}_u X_v. \quad (6.93)$$

We now derive the dynamical equations of motion. Carrying out the analogous derivation as we did in Sec. 5.2.2, we find that the kinetic energy density per unit material area is

$$\mathcal{K} = \frac{1}{2}\rho_0|\vec{V}|^2 + \frac{1}{2}\vec{\Omega}^T \mathbb{I} \vec{\Omega} \quad (6.94)$$

where $\rho_0 = L_0^1 \rho_0^V$ and \mathbb{I} is the moment-of-inertia of the material frame per unit material area.

From Eq. 6.12 we find the dynamical equations of motion

$$\mathcal{D}_t^* S = \mathcal{D}_u^* Q^u + \mathcal{D}_v^* Q^v + T \quad (6.95a)$$

$$n_\alpha Q^\alpha = 0, \quad \text{on } \partial M. \quad (6.95b)$$

where we sum over $\alpha = u, v$ and ad^* was given in Eq. 5.81, and where

$$S = \frac{\partial \mathcal{K}}{\partial N} = \{\vec{P}; \vec{L}\}^* \quad (6.96a)$$

$$Q^u = \{\vec{F}^u; \vec{M}^u\}^* \quad (6.96b)$$

$$Q^v = \{\vec{F}^v; \vec{M}^v\}^* \quad (6.96c)$$

$$T = \{\vec{f}; \vec{m}\}^* \quad (6.96d)$$

and $\vec{P} = \rho_0 \vec{V}$ and $\vec{L} = \mathbb{I} \vec{\Omega}$. Substituting Eq. 6.90 and Eq. 6.96 into Eq. 6.95 we get

$$D_t \vec{P} = D_\alpha \vec{F}^\alpha + \vec{f} \quad (6.97a)$$

$$D_t \vec{L} = D_\alpha \vec{M}^\alpha + \vec{\theta}_\alpha \times \vec{F}^\alpha + \vec{m} \quad (6.97b)$$

$$n_\alpha \vec{F}^\alpha = n_\alpha \vec{M}^\alpha = 0, \quad \text{on } \partial M \quad (6.97c)$$

where we sum over repeated indices of $\alpha = u, v$, and is consistent with the dynamical equations of motion found in the literature [181]. Repeating the same dimensional arguments as in Sec. 6.3.1, we can conclude that \vec{P} and \vec{L} have units of momentum and angular momentum per unit material area respectively, and \vec{F}^α and \vec{M}^α force and moment per unit material length respectively. The body force \vec{f} and \vec{m} have units of force and moment per unit material area respectively.

Note that the component L_1 of the angular momentum of the material frame is unphysical if the Cosserat surface is seen as an approximate model for a slender shell. For conservative dynamics, this means that \mathcal{U} should not couple to π_1 , which is the rate-of-twist of the material frame around \mathbf{e}_1 . In general, this entails that $M_1^\alpha = m_1^\alpha = 0$.

In the above we have considered a rectangular material base space M . However, in general for *open* Cosserat surfaces the material base space can be any bounded and compact subset $M \subset \mathbb{R}^2$. Furthermore, *closed* Cosserat surfaces may have $M = S^2$, in which case the material base space no longer admits global coordinates. The kinodynamical equations of motion above still apply in the case of a closed surface, but they must then be simultaneously and consistently simulated over multiple charts that cover M .

6.3.3 Cosserat rods on 2-spheres

We consider the constitutive dynamics of a generalised Cosserat rod, lying on the 2-dimensional surface of a sphere $S^2 = \{\mathbf{x} \in \mathbb{E}^3 \mid |\mathbf{x}| = r\}$ where r is the radius of the sphere. Some recent applications of such systems can be found in [249, 250]. The external configuration space of the rod is thus $S^2 = SO(3)/SO(2)$, and the internal configuration space is $SO(2)$. Let $\mathbf{r}(t, u) = r \mathbf{e}_0(t, u)$ denote the center-line of the rod, where $\mathbf{n}(t, u) \in \mathbb{E}^3$ is a unit-vector, and $u \in [0, L_0]$, $t \in [0, T]$. The material frame $E = (\mathbf{e}_1(t, u), \mathbf{e}_2(t, u))$ of

the rod are two orthonormal vectors that are tangent to S^2 at \mathbf{r} , and can be seen as an element of $SO(2)$. We also have that the triad $(\mathbf{n} \ \mathbf{e}_1 \ \mathbf{e}_2)$ are mutually orthogonal, and $\mathbf{n}(t, u)$ is normal to S^1 at $\mathbf{r}(t, u)$ and $(\mathbf{e}_1 \ \mathbf{e}_2)$ forms a basis for $T_{\mathbf{r}(t,u)}S^2$. Furthermore, as the triad forms an orthogonal matrix with unit determinant we can identify it as an element of $SO(3)$, and will thus write $\Phi = (\mathbf{n} \ \mathbf{e}_1 \ \mathbf{e}_2)$. The configuration space of the Cosserat rod on the sphere is thus $G = SO(3)$, with kinematic base space $W = [0, T] \times [0, L_0]$ which admits a global basis. As in Ch. 5 we will distinguish between vectors $\mathbf{v} \in \mathbb{E}^3$ in the fixed frame and vectors $\vec{v} \in \mathbb{R}^3$ in the moving frame, and relate the two as $\vec{v} = \Phi^T \mathbf{v}$. We shall write the components of vectors in the moving frame as $\vec{v} = (v_n, v_1, v_2)$.

We shall write the Maurer-Cartan form as

$$\xi = \hat{N}dt + \hat{X}du \quad (6.98)$$

where $\hat{N}(t, u), \hat{X}(t, u) \in \mathfrak{so}(3)$, and we write

$$\vec{N} = (\Omega_n, -V_2/r, V_1/r)^T \quad (6.99a)$$

$$\vec{X} = (\pi_n, -\theta_2/r, \theta_1/r)^T \quad (6.99b)$$

and

$$\vec{V} = (0, V_1, V_2)^T \quad (6.100a)$$

$$\vec{\theta} = (0, \theta_1, \theta_2)^T \quad (6.100b)$$

$$\vec{\Omega} = (\Omega_n, 0, 0)^T \quad (6.100c)$$

$$\vec{\pi} = (\pi_n, 0, 0)^T. \quad (6.100d)$$

From $\Phi^{-1}d\Phi = \xi$ we find

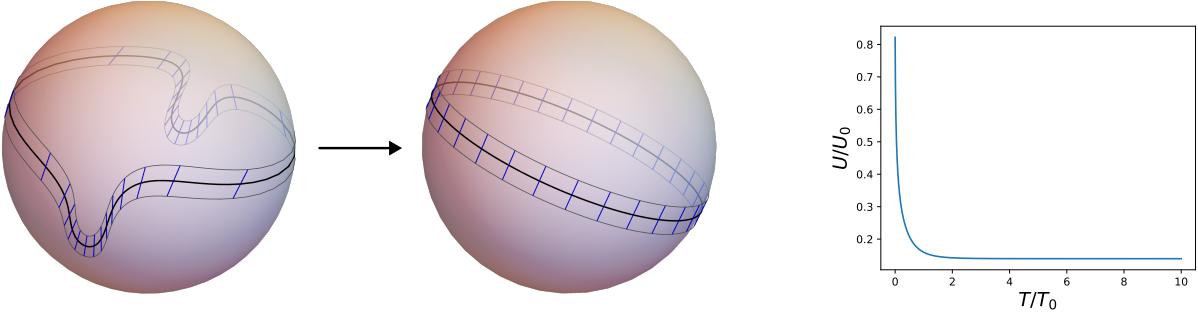
$$d\mathbf{r} = r d\mathbf{n} = \mathbf{e}_i V_i dt + \mathbf{e}_i \theta_i du \quad (6.101a)$$

$$= \mathbf{V} dt + \boldsymbol{\theta} du$$

$$d\mathbf{e}_1 = (\Omega_n \mathbf{e}_2 - (V_1/r) \mathbf{n}) dt + (\pi_n \mathbf{e}_2 - (\theta_1/r) \mathbf{n}) du \quad (6.101b)$$

$$d\mathbf{e}_2 = (-\Omega_n \mathbf{e}_1 - (V_2/r) \mathbf{n}) dt + (-\pi_n \mathbf{e}_1 - (\theta_2/r) \mathbf{n}) du. \quad (6.101c)$$

We can identify $\mathbf{V}(t, u) \in T_{\mathbf{r}(t,u)}S^2$ as the velocity of the center-line at $\mathbf{r}(t, u)$ and likewise $\boldsymbol{\theta}(t, u) \in T_{\mathbf{r}(t,u)}S^2$ as the rate-of-change of the center-line along the material coordinate u . We can also identify $\Omega_n(t, u)$ and $\pi_n(t, u)$ as the angular velocity and angular rate-of-rotation of the material frame around \mathbf{n} , and thus $\Omega(t, u), \pi(t, u) \in \mathfrak{so}(2)$.



(a) Overdamped Cosserat rod relaxing to a ground-state

(b) Potential energy

Figure 6.3: (a) Overdamped Cosserat rod on 2-sphere relaxing from a deformed initial configuration (left) to a ground-state (right). Solid black lines are the rod center-lines, and the blue lines are the directors \mathbf{e}_2 . (b) Potential energy of the Cosserat rod.

From Eq. 6.5 we can find the kinematic equations of motion

$$\begin{aligned}\dot{\vec{X}} &= \mathcal{D}_u \vec{N} \\ &= \vec{N}' + \vec{X} \times \vec{N}\end{aligned}\tag{6.102}$$

using $\overrightarrow{[\hat{X}, \hat{N}]} = \vec{X} \times \vec{N}$. We define the Lagrangian

$$\mathcal{L}(d\Phi) = \mathcal{K}(\dot{\Phi}) - \mathcal{U}(\Phi')\tag{6.103a}$$

$$\mathcal{K} = \frac{1}{2}\rho_0|\mathbf{V}|^2 + \frac{1}{2}\mathbb{I}_n\Omega_n^2\tag{6.103b}$$

where \mathcal{U} is constitutive potential energy, ρ_0 is the mass density per unit material length and $\mathbb{I}_n \in \mathbb{R}$ is the moment-of-inertia per unit material length of the material frame. In its reduced form, the Lagrangian is

$$\ell(\xi) = \mathcal{K}(N) - \mathcal{U}(X)\tag{6.104a}$$

$$\mathcal{K} = \frac{1}{2}\rho_0|\vec{V}|^2 + \frac{1}{2}\mathbb{I}_n\Omega_n^2.\tag{6.104b}$$

The conjugate moment and stress are $\vec{P} = \frac{\partial \ell}{\partial \vec{N}}$ and $\vec{Q} = -\frac{\partial \ell}{\partial \vec{X}}$ respectively, and we get

$$\vec{P} = (L_n, p_2 r, -p_1 r)^T\tag{6.105a}$$

$$\vec{Q} = (M_n, F_2 r, -F_1 r)^T\tag{6.105b}$$

where $\vec{F} = (0, F_1, F_2)^T$ and $\vec{p} = (0, \rho_0 V_1, \rho_0 V_2)^T$ is the force on the center-line and its linear momentum density per unit material length respectively, and where $\vec{M} = (M_n, 0, 0)$ and $\vec{L} = (\mathbb{I}_n\Omega_n, 0, 0)^T$ are the moment on the material frame and its angular momentum per unit material length respectively. From Eq. 6.12 we find the dynamical equations of

motion

$$(\partial_t + \hat{N})\vec{S} = (\partial_u + \hat{X})\vec{Q} + \vec{T} \quad (6.106a)$$

$$Q_\alpha = 0, \quad u = 0, L_0. \quad (6.106b)$$

where we used $[\text{ad}_X] = \hat{X}$ and $[\text{ad}_X^*] = -[\text{ad}_X]^T$, and where \vec{T} is a generalised body force. In the absence of body forces, and using the definitions of S, Q, N and X , we get

$$\dot{\vec{L}} = M' + \vec{\theta} \times \vec{F} \quad (6.107a)$$

$$\dot{\vec{P}} = D_u \vec{F} - c\vec{\Omega} \times \vec{V} + \frac{1}{r^2} \vec{M} \times \vec{\theta} \quad (6.107b)$$

where $c = \rho_0 + \mathbb{I}_n/r^2$. $\vec{\theta} \times \vec{F}$ is the moment exerted on the material frame due to the force, $\frac{1}{r^2} \vec{M} \times \vec{\theta}$ is the force exerted on the center-line due to the moment, and $-c\vec{\Omega} \times \vec{V}$ can be seen as a Coriolis force that arises due to the rotating frame. Equations 6.102 and 6.107a together form a set of equations that completely determine the kinodynamics of the system.

We will now consider the example of ‘overdamped’ Cosserat rod dynamics on the 2-sphere. We define the frictional body force $\vec{T} = \gamma \vec{N}$, where $\gamma \in \mathbb{R}^{3 \times 3}$ is a symmetric and positive-definite matrix, and the constitutive potential

$$\mathcal{U} = \frac{1}{2} \epsilon (\vec{X} - \vec{X}_0)^T \mathbf{K} (\vec{X} - \vec{X}_0) \quad (6.108)$$

where $\vec{X}_0 = (0 \ 1 \ 0)^T$ and $\mathbf{K} \in \mathbb{R}^{3 \times 3}$ is a positive definite matrix. We take the overdamped limit, eliminating the inertial degrees of freedom $\vec{P} \approx 0$, to get

$$\dot{\vec{X}} = \vec{N}' + \vec{X} \times \vec{N} \quad (6.109a)$$

$$\gamma \vec{N} = (\partial_u + \hat{X})\vec{Q}. \quad (6.109b)$$

Fig. 6.3 shows the result of a simulation of Eq. 6.109, depicting the relaxation of an initial deformed state to a ground-state that minimises Eq. 6.108.

6.3.4 Relativistic Cosserat rods

Here we derive the kinematic equations of relativistic motion of a Cosserat rod. We work in units where the speed-of-light constant is set to unity $c = 1$. The Minkowski vector space $\mathbb{M}^{1,3}$ is the vector space \mathbb{R}^4 equipped with the *Minkowski inner product* $\langle \cdot, \cdot \rangle_M : \mathbb{M}^{1,3} \times \mathbb{M}^{1,3} \rightarrow \mathbb{R}$ with signature $(1, 3)$. In other words, given some basis $B = (\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ for \mathbb{R}^4 , the *Minkowski metric*

$$\eta_{ij} = \langle \mathbf{b}_i, \mathbf{b}_j \rangle_M \quad (6.110)$$

has 1 negative eigenvalue and 3 positive eigenvalues. Henceforth we will assume that the basis is defined such that

$$\eta = \text{diag}(-1, 1, 1, 1). \quad (6.111)$$

Any basis that diagonalises η as in Eq. 6.111 will be called an *orthonormal* basis. A $\mathbf{v} \in \mathbb{M}^{1,3}$ is known as *time-like* if $\langle \mathbf{v}, \mathbf{v} \rangle_M < 0$, *space-like* if $\langle \mathbf{v}, \mathbf{v} \rangle_M > 0$ and *light-like* if $\langle \mathbf{v}, \mathbf{v} \rangle_M = 0$. We can thus identify \mathbf{b}_0 as the time-like direction in this basis, and $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ as the space-like directions.

The *space-time coordinates* $\mathbf{r}(\tau)$ of an observer is a function $\mathbf{r} : \mathbb{R} \rightarrow \mathbb{M}^{1,3}$ where τ is the time measured by clocks co-moving with the observer, known as the *proper time*. The *4-velocity* of the observer is given by the time-like vector $\mathbf{U} = \partial_\tau \mathbf{r}$. The *inertial frame* of the observer at proper time τ is an orthonormal basis $E(\tau) = (\mathbf{e}_0(\tau), \mathbf{e}_1(\tau), \mathbf{e}_2(\tau), \mathbf{e}_3(\tau))$ such that

$$\vec{U} = E^{-1} \mathbf{U} = (1 \ 0 \ 0 \ 0)^T. \quad (6.112)$$

Intuitively, this corresponds to the fact that an observer is always stationary in its own co-moving inertial reference frame. Such a basis can be found by setting $\mathbf{e}_0 = \mathbf{U}$, and the remaining three basis elements $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ specify the spatial orientation of the observer. Vectors $\vec{v} = E^{-1} \mathbf{v}$ are thus expressed in the inertial frame E . We shall use the notation \tilde{v} to refer to the spatial components of \vec{v} , such that $\vec{v} = (v_0 \ \tilde{v}^T)^T$.

Any two inertial frames E_1 and E_2 can be related by a *Lorentz transformation*

$$E_2 = \Lambda E_1 \quad (6.113)$$

where $\Lambda \in SO(1, 3)$, and where $SO(1, 3)$ is the *Lorentz group* on $\mathbb{M}^{1,3}$ and is defined as

$$SO(1, 3) = \{\Lambda \in \mathbb{R}^{4 \times 4} \mid \langle \Lambda \mathbf{v}, \Lambda \mathbf{v} \rangle_M = \langle \mathbf{v}, \mathbf{v} \rangle_M \ \forall \mathbf{v} \in \mathbb{M}^{3,1}\}. \quad (6.114)$$

which is the group of rotations in space and *Lorentz boosts*. The Lorentz group is thus the set of linear transformations that preserves the Minkowski inner product. Combined with the group of translations on $\mathbb{M}^{1,3}$, we have the *Poincaré* group

$$M(1, 3) = \left\{ \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{t} & \Lambda \end{pmatrix} \in \mathbb{R}^{5 \times 5} \mid \mathbf{t} \in \mathbb{M}^{1,3}, \ \Lambda \in SO(3, 1) \right\} \quad (6.115)$$

of space-time translations and rotations.

Now consider a continuous ‘string’ of inertial observers with space-time coordinates $\mathbf{r}(\tau, u)$ and inertial frames $E(\tau, u)$, parametrised by u . If $\mathbf{e}_0(\tau, u) = \partial_\tau \mathbf{U}(\tau, u)$ then this can be considered a *relativistic* Cosserat rod, where $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ is here the material frame of the rod. In order to avoid confusing using the word ‘frame’ in reference to both material and inertial frames, we will henceforth refer to the former as the Cosserat cross-section.

Given a reference frame B and origin $\mathbf{0} \in \mathbb{M}^{1,3}$, we can write the configuration of the Cosserat rod as

$$\mathcal{F} = \mathcal{F}_0 \Phi \quad (6.116)$$

where

$$\mathcal{F} = (\mathbf{r}; E) = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{r} & E \end{pmatrix} \quad (6.117)$$

and $\mathcal{F}_0 = (\mathbf{0}; B)$ and $\Phi = (B^{-1}\mathbf{r}; B^{-1}\Lambda B)$, and where Λ satisfies $E = \Lambda B$. As in Sec. 5.2.1, we now simplify our notation by letting $B = \mathbb{1}_{4 \times 4}$, such that $\Lambda = E \in SO(1, 3)$ and we can thus write $\Phi = \mathcal{F}$.

We write the Maurer-Cartan form as

$$\xi = N d\tau + X du \quad (6.118)$$

where

$$N = \{\vec{U}; O\} = \begin{pmatrix} 0 & \vec{0}^T \\ \vec{U} & O \end{pmatrix} \quad (6.119a)$$

$$X = \{\vec{\theta}; \Xi\} = \begin{pmatrix} 0 & \vec{0}^T \\ \vec{\theta} & \Xi \end{pmatrix} \quad (6.119b)$$

and $N(\tau, u), X(\tau, u) \in \mathfrak{m}(1, 3)$ and $O, \Xi \in \mathfrak{so}(1, 3)$. From $\Phi^{-1}d\Phi = \xi$ we find

$$d\mathbf{r} = \mathbf{U} d\tau + \boldsymbol{\theta} du \quad (6.120a)$$

$$= \mathbf{e}_0 d\tau + \boldsymbol{\theta} du$$

$$d\mathbf{e}_\beta = \mathbf{e}_\alpha O_{\alpha\beta} d\tau + \mathbf{e}_\alpha \Xi_{\alpha\beta} du, \quad \alpha, \beta = 0, 1, 2, 3 \quad (6.120b)$$

where we used $\mathbf{e}_0 = \mathbf{U}$, which should be seen as a kinematic restriction on the Lie group-valued Φ (although note that it is not a kinematic restriction from the physical perspective, as real inertial observers always are always stationary with respect to themselves). Because we have parametrised the Cosserat rod with respect to the co-moving inertial frames, we have that $\vec{U}(\tau, u) = (1 \ 0 \ 0 \ 0)^T$. Therefore the kinematic movement of the rod is not specified in \vec{U} as in Sec. 5.2, but must instead be encoded in O .

The Lie algebra element $O \in \mathfrak{so}(1, 3)$ can be written as

$$O = \begin{pmatrix} 0 & \tilde{a}^T \\ \tilde{a} & \hat{\tilde{\Omega}} \end{pmatrix} \quad (6.121)$$

where $\tilde{a} = (a_1, a_2, a_3)$ and $\hat{\tilde{\Omega}} \in \mathfrak{so}(3)$. Henceforth the tilde will designate spatial 3-vectors.

To interpret these quantities, we first note that

$$\partial_\tau \mathbf{e}_i = \mathbf{e}_i \hat{\tilde{\Omega}}_{ij}, \quad i = 1, 2, 3 \quad (6.122)$$

from which we identify that $\tilde{\Omega}(\tau, u)$ is the angular velocity of the Cosserat cross-section at u in its inertial frame. Secondly, we can compute the 4-acceleration $\mathbf{a} = \partial_\tau^2 \mathbf{r}$ as

$$\mathbf{a} = \partial_\tau \mathbf{e}_i = \mathbf{e}_i a_i, \quad i = 1, 2, 3. \quad (6.123)$$

such that $\vec{a} = E^{-1} \mathbf{a} = (0 \ \tilde{a}^T)^T$, which is the correct expression for the co-moving 4-acceleration in special relativistic kinematics [251, p. 99]. We thus see that $\tilde{a}(\tau, u)$ is the acceleration of the Cosserat cross-section at u in its co-moving inertial frame. Therefore the kinematics of the relativistic Cosserat rod is specified by the spatial acceleration \tilde{a} of the center-line and the angular velocity $\tilde{\Omega}$ of the cross-section. This stands in contrast to the non-relativistic Cosserat rod, where we instead specify the velocity of the center-line. We can understand this difference by noting that velocity is *itself* a kinematic degree of freedom in special relativity, in addition to position and orientation. Only the latter two are kinematic degrees of freedom in non-relativistic systems. Mathematically, we have that the relative velocity of an inertial observer with frame E_1 with respect to another observer with frame E_2 is encoded in the Lorentz transformation that relates the two, as in Eq. 6.113. This is therefore the reason why we must specify the *acceleration* of the frame, as opposed to its *velocity*, in the kinematics.

The kinematic equations of motion are given by Eq. 6.5, from which we find

$$\partial_\tau \vec{\theta} = -O \vec{\theta} + \Xi \vec{U} \quad (6.124a)$$

$$\partial_\tau \Xi = \partial_u O + [\Xi, O]. \quad (6.124b)$$

If we let

$$O = \begin{pmatrix} 0 & \tilde{t}^T \\ \tilde{t} & \hat{\tilde{\pi}} \end{pmatrix} \quad (6.125)$$

then the kinematic equations of motion can be written as

$$\partial_\tau \theta_0 = a_i \theta_i \quad (6.126a)$$

$$\tilde{D}_\tau \tilde{\theta} = -\theta_0 \tilde{a} + \tilde{t} \quad (6.126b)$$

$$\tilde{D}_\tau \tilde{t} = \tilde{D}_u \tilde{a} \quad (6.126c)$$

$$\partial_\tau \tilde{\pi} = \tilde{D}_u \tilde{\Omega} \quad (6.126d)$$

where $i = 1, 2, 3$ and $\tilde{D}_\tau = \partial_\tau + \hat{\tilde{\Omega}}$ and $\tilde{D}_u = \partial_u + \hat{\tilde{\pi}}$.

6.3.5 $O(n)$ -NLSM field theory

Here we give an example of the GGCT as applied to a field theory. We consider a field $\vec{p} : W \rightarrow S^N \subset \mathbb{R}^N$, where W is the base space and $n\vec{p}$ is a unit-vector in \mathbb{R}^N for all $p \in W$. We call $S^N \subset \mathbb{R}^N$ the *target manifold* of the field theory. We let $W = \mathbb{M}^{1,d}$, which is the $(d+1)$ -dimensional Minkowski space, although what follows is easily generalisable to Euclidean space or any arbitrary base space. We define coordinates $(t, u^1, u^2, \dots, u^d)$ on $\mathbb{M}^{1,d}$, and we will write partial derivatives as ∂_γ , $\gamma = 0, 1, \dots, d$, where $\partial_0 = \frac{\partial}{\partial t}$ and $\partial_\alpha = \frac{\partial}{\partial u^\alpha}$, $\alpha = 1, \dots, d$.

The $O(n)$ non-linear σ model (NLSM) [244] is defined by the Lagrangian density

$$\mathcal{L} = \frac{1}{2} \eta^{\gamma\kappa} (\partial_\gamma \vec{p})^T (\partial_\kappa \vec{p}) \quad (6.127)$$

which is expressed in some local coordinate chart of $\mathbb{M}^{1,d}$, where η is the Minkowski metric on $\mathbb{M}^{1,d}$ with signature $(1, d)$.

Let the $SO(n)$ -valued field $\Phi : \mathbb{M}^{1,d} \rightarrow SO(3)$ satisfy $\vec{p} = \Phi \vec{p}_0$ where $\vec{p}_0 \in S^N$ is some constant reference vector. We can then rewrite Eq. 6.127 as

$$\mathcal{L} = \frac{1}{2} \eta^{\gamma\kappa} (\partial_\gamma \Phi \vec{p}_0)^T (\partial_\kappa \Phi \vec{p}_0) \quad (6.128)$$

Now, we have

$$\begin{aligned} (\partial_\gamma \Phi \vec{p}_0)^T (\partial_\kappa \Phi \vec{p}_0) &= (\Phi^{-1} \partial_\gamma \Phi \vec{p}_0)^T (\Phi^{-1} \partial_\kappa \Phi \vec{p}_0) \\ &= (\hat{Z}_\gamma \vec{p}_0)^T (\hat{Z}_\kappa \vec{p}_0) \end{aligned} \quad (6.129)$$

for all $\gamma, \kappa = 1, \dots, d$, where we have defined $\hat{Z}_\gamma = \Phi^{-1} \partial_\gamma \Phi \in \mathfrak{so}(n)$. As before we will be making use of the hat-map to convert anti-symmetric 3×3 -matrices \hat{Z}_γ to 3-vectors \vec{Z}_γ . Let us now assume that the coordinates are defined such that $\eta = \text{diag}(-1, 1, \dots, 1)$, where the speed-of-light constant has been set to unity $c = 1$. The Lagrangian then has a reduced form

$$\ell(\xi) = -\frac{1}{2} \vec{N}^T \mathbf{N}_0 \vec{N} + \frac{1}{2} \sum_{\alpha=1}^d \vec{X}_\alpha^T \mathbf{N}_0 \vec{X}_\alpha \quad (6.130)$$

where $\vec{N} = \vec{Z}_0$ and $\vec{X}_\alpha = \vec{Z}_\alpha$, $\alpha = 1, \dots, d$ and $\mathbf{N}_0 = \hat{n}_0^T \hat{n}_0$ and we have used $|\hat{Z}_\gamma \vec{p}_0| = |\vec{Z}_\gamma \times \vec{p}_0| = |\hat{n}_0 \vec{Z}_\gamma|$.

We can then readily apply the GGCT to the system, to get the field equations of motion

$$\partial_t \vec{X}_\alpha = (\partial_\alpha + \hat{X}_\alpha) N \quad (6.131a)$$

$$(\partial_t + \hat{N}) \vec{S} = \sum_{\alpha=1}^d (\partial_\alpha + \hat{X}_\alpha) \vec{Q}^\alpha \quad (6.131b)$$

where $\vec{p} = \vec{Z}_0$, $\vec{X}_\alpha = \vec{Z}_\alpha$, $\alpha = 1, \dots, d$, $S = \frac{\partial \ell}{\partial N}$ and $Q^\alpha = \frac{\partial \ell}{\partial X_\alpha}$. The spatial reconstruction fields X_α must satisfy the spatial integrability conditions

$$\partial_\beta X_\alpha = \mathcal{D}_\alpha X_\beta \quad (6.132)$$

at all times t , where $\alpha = 1, \dots, d-1$ and $\beta = \alpha + 1, \dots, d$.

Note that only two of the three components of S are independently determined by Eq. 6.131b, as N_0 is only a rank 2 matrix. To see this let $\vec{p}_0 = (1 \ 0 \ 0)^T$ such that $N_0 = \text{diag}(0, 1, 1)$. We then have that $\vec{S} = (0, N_2, N_3)$ and $Q^\alpha = (0, X_2, X_3)$. This reflects the fact that rotations of \vec{p} around its own axis is a gauge freedom in our formulation of the $O(n)$ non-linear σ model.

6.4 Conclusion

In this chapter we presented a framework to formulate and study the kinodynamics of generalised Cosserat systems, which are continuum bodies with configuration spaces that are Lie groups G , or homogeneous spaces \mathbb{X} . The result was a generalised geometric theory of Cosserat rods (GGCT), which we built by extrapolating from the geometric features of the Cosserat rod derivation, presented in Ch. 5. Specifically, our framework can be seen as a generalised mechanics of microstructured continuum bodies, and from it we recovered the kinodynamics of Cosserat rods, sheets and bodies, as well as that of classical continuum mechanics. However, we also showed that its application as a theoretical framework extended beyond this interpretation, as we further applied it to Cosserat rods in Minkowski space, to a non-linear σ field theory and Cosserat rods on 2-spheres. In so doing we aimed to show the ease-of-use, as well as the wide applicability of our approach.

In full generality, the broad class of system we considered were G - and \mathbb{X} -valued fields on topological manifolds M . We called this field the spatio-temporal configuration Φ . However, our equations of motions were not formulated in terms of Φ . Rather, the field configuration was encoded in terms of Lie algebraic quantities X_α , which - via the exponential map - integrate to Φ . From this perspective, the GGCT is a framework using which we can derive Lie algebraic kinodynamical equations of motion of G - and \mathbb{X} -field theories.

One of the main motivations for formulating the kinodynamics of a system on the corresponding Lie algebra of its configuration space is that this naturally lends itself to studying the system in terms of its intrinsic geometry. As opposed to finite-dimensional systems, which do not admit the notion of a constitutive force, for any realistic continuum system the intrinsic geometry must couple to the dynamics. That is, constitutive forces are functions of the derivative $d\Phi \in TG$, which is a vector field on G , rather than Φ itself.

Now, the tangent bundle TG of any Lie group admits a trivialisation to the tangent space at the identity, the Lie algebra. The cornerstone of the GGCT was to use this trivialisation to express all vector and covector fields as Lie algebra-valued fields. In a reductive, albeit simplified, manner of speaking: This allowed us to express kinodynamics directly in terms of $X_\alpha \sim d\Phi$, thus avoiding having to take the derivative $d\Phi$.

Future work may involve several parallel lines of research. In this chapter we focused primarily on constructing the framework itself, and we ‘applied’ it in the sense that we used the framework to show that it yielded the correct kinodynamics of a variety of known (and new) systems. In future research we aim to add to the literature of active continua [171, 172, 173, 174], in particular we aim to study activity on surfaces [39, 168, 227, 252]. The analytical and numerical challenges of such systems will be a testing ground for the formalism developed in this chapter.

Notably, and in contrast to the first part of this thesis, we have only considered deterministic dynamics here. One major avenue for theoretical research is to incorporate stochasticity into the GGCT. Developments in this direction have been made as regards to filaments and polymer chains [253, 254, 255, 256]. In the spirit of the programme outlined in this text, we would characterise and carefully define stochasticity on Lie group continua [257]. In particular, we would aim to generalise the notion of detailed balance to general G - and \mathbb{X} -configured continuum system. For example, it may be possible to accomplish this by reformulating the dynamics from the Lagrangian to the Hamiltonian perspective, of which the former was used in this text. Then, assuming that the statistics of the system configuration is that of a Boltzmann distribution, we can derive the conditions for detailed balance. This is analogous to how detailed balance is derived in stochastic field theories [138]. Furthermore, we note that if our ambitions of constructing MCMC methods to sample field theories come to fruition, then these could be used to sample the configurations of the stochastic continua described here.

Finally, there is substantial work to be done in constructing geometric numerical integrators adapted to G - and \mathbb{X} -configured continua. There is already an extensive literature on such integrators for finite-dimensional systems [258, 259, 260, 261]. In the following chapter we will derive the first instance of a geometric integrator defined on the general continuum we have defined in this chapter.

Chapter 7

Geometric numerical integrators

In numerics, the general class of systems considered in Ch. 6 are simulated by integrating the kinodynamic equations of motion, which were stated in full in Eq. 6.20. One of the simplest methods to use is the Forward-Euler integrator (FEI), wherein the equations are linearised by means of a uniform discretisation in time. However, the geometric intricacies of Eq. 6.20 are lost if such a scheme is applied. In particular, the adjoint action of the Lie group G is prominent in the equations of motion, where G is either the symmetry group of the configuration space or the configuration space itself. In this chapter we develop numerical integrators that are built around the properties of G , which are constructed so that solutions retain the qualitative geometric features of their equations of motion. As we show in Sec. 7.4, the resulting *geometric* numerical integrators preserve the spatial integrability conditions Eq. 6.21a. There is an extended literature on similar integrators to those that we derive here for finite-dimensional Lie group-valued Hamiltonian systems [262, 263, 264].

The kinodynamic equations of motion Eq. 6.20 are all expressed in terms of Lie algebra-valued, or dual Lie algebra-valued, quantities. Using the fundamental representation of the Lie algebra and its dual, we get equations in terms of the \mathbb{R} -valued components of their matrix representations. Therefore we will begin by studying the most general form in which our equations of motion appear

$$\partial_t x = F + Ax \quad (7.1)$$

where $x = x(t, \mathbf{u}) \in \mathbb{R}^n$ is the state of the system, and F is a \mathbb{R}^n -valued drift-field and A a $\mathbb{R}^{n \times n}$ -valued linear operator acting on the components of x . In general, F and A can be dependent on x as well as its derivatives $\partial_\alpha x, \partial_\alpha^2 x$, including higher-order derivatives. Given an initial condition $x(t_0) = x_0$ and $F_0 = F(x_0, \partial_\alpha x|_{x=x_0}, \dots)$ and $A_0 = A(x_0, \partial_\alpha x|_{x=x_0}, \dots)$, then we can approximate $x(t_0 + \Delta t)$ as

$$x(t_0 + \Delta t, \mathbf{u}) \approx F_0 \Delta t + A_0 x_0 \Delta t. \quad (7.2)$$

which is an approximate short-time propagator. This is the Forward-Euler integration scheme, using which we can numerically approximate solutions to Eq. 7.1 on a uniform time-grid $t_i = i\Delta t$. Implicitly we have assumed that f and the product Ax_0 are constant in the time-interval $[t_0, t_0 + \Delta t]$ in Eq. 7.2. We now relax the latter assumption, approximating Eq. 7.1 as

$$\partial_t x \approx F_0 + A_0 x, \quad t \in [t_0, t_0 + \Delta t]. \quad (7.3)$$

where $x(t_0) = x_0$. The right-hand side can be solved analytically [265], such that we get the alternative short-time propagator

$$x(t_0 + \Delta t, \mathbf{u}) \approx e^{\Delta t A_0} x_0 + \left(\int_0^{\Delta t} e^{(\Delta t - \Delta t') A_0} d\Delta t' \right) F_0, \quad (7.4)$$

which provides an alternative to the FEI Eq. 7.3.

Now, let \mathfrak{g} be a Lie algebra with representation $\rho : \mathfrak{g} \rightarrow \mathfrak{gl}(V)$, where V is an n -dimensional vector space and $\mathfrak{gl}(V)$ is the Lie algebra of the general linear group. If we recast Eq. 7.1 such that A is an element of the given Lie algebra representation (that is, $A \in \rho(\mathfrak{g})$), then x takes values in the vector space V that the Lie algebra is acting on. Furthermore, the drift-field is now a function $F : V \rightarrow V$, and should be considered to take values in $\mathfrak{t}(V)$, the Lie algebra of the *translation group*¹ $T(V)$ acting on V . The right-hand side of Eq. 7.1 can thus be seen as a combination of a translation $F \in \mathfrak{t}(V)$ and the \mathfrak{g} -action $A \in \rho(\mathfrak{g})$ on the V -valued x . Therefore, the linear map

$$x \mapsto F + Ax \quad (7.5)$$

can *itself* be seen as a representation of the product Lie algebra $\mathfrak{t}(V) \times \mathfrak{g}$. This can be seen explicitly by rewriting Eq. 7.1 as

$$\partial_t \begin{pmatrix} 1 \\ x \end{pmatrix} = M \begin{pmatrix} 1 \\ x \end{pmatrix} \quad (7.6)$$

where

$$M = \begin{pmatrix} 0 & 0 \\ F & A \end{pmatrix}, \quad (7.7)$$

where we identify M as a representation of $\mathfrak{t} \times \mathfrak{g}$.

Now, if we repeat the same derivation we did previously to get a short-time propagator

¹Note that as the translation group is abelian, we have that it is isomorphic with its Lie algebra $\mathfrak{t}(V) \cong T(V)$.

for Eq. 7.6. Let $x(t_0) = x_0$ and $M_0 = M|_{x=x_0}$, we then have

$$\begin{pmatrix} 1 \\ x(t + \Delta t), \mathbf{u} \end{pmatrix} \approx e^{\Delta t M_0} \begin{pmatrix} 1 \\ x_0 \end{pmatrix}. \quad (7.8)$$

Recall that the exponential map of a Lie algebra representation yields a representation of the corresponding Lie group. Therefore, therefore we have that $e^{\Delta t M_0}$ is a representation of the Lie group product $T \times G$. Now, note that Eq. 7.8 must be equivalent Eq. 7.4. This tells us that the right-hand side of Eq. 7.1 can be seen as the infinitesimal action of the Lie algebra $\mathfrak{t}(V) \times \mathfrak{g}$ on the V -valued x , via the given representation. Eq. 7.4 thus ensures, using the exponential map $\exp : \mathfrak{t}(V) \times \mathfrak{g} \rightarrow T \times G$, that we properly integrate the infinitesimal action from the Lie algebra to the corresponding Lie group $T(V) \times G$. In this setting, the integrator Eq. 7.4 is known as a *Lie group integrator* (LGI) [258, 259, 260, 261]. Both the FEI and the LGI will suffer numerical errors. However, the errors of the former will accrue in such a way that the ‘effective’ action of F and A on x does not exponentiate to a well-defined action on the Lie group-level. On the other hand, the LGI is defined so that errors accrue *within the Lie group itself* [260].

In Sec. 7.1 we will use the results above to derive Lie group integrators for the general class of G/H -configured systems considered in Ch. 6. We will then present results for the specific cases of $SO(3)$ - and $SE(3)$ -valued configuration spaces in Sec. 7.2 and Sec. 7.3 respectively. The results of these latter two sections also apply for systems with a configuration space that is a homogeneous space for which $SO(3)$ or $SE(3)$ is the symmetry group. Sec. 7.4 we show some numerical results for the Cosserat rod and Cosserat sheet, simulated using the geometric integrators.

7.1 Geometric integrators for general configuration spaces

We consider a generalised Cosserat system, with a d -dimensional material base space M and configuration space G/H , where G is an n -dimensional Lie group with Lie algebra \mathfrak{g} and $H \subset G$ is a Lie subgroup. The kinodynamic equations of motion of the system are given in Eq. 6.20, which can be put into the form of Eq. 7.1 by setting

$$\partial_t X_\alpha = F_\alpha + A X_\alpha \quad (7.9a)$$

$$\partial_t S = H + B S \quad (7.9b)$$

where $\alpha = 1, \dots, d$ and

$$F_\alpha = \partial_\alpha N \quad (7.10a)$$

$$H = \mathcal{D}_\alpha^* Q^\alpha + T \quad (7.10b)$$

and $A = -\text{ad}_N$, $B = -\text{ad}_N^*$, where we used $\text{ad}_X N = -\text{ad}_N X$. We assume that all quantities are expressed in a non-dimensionalised form. Now, as per the discussion in the introduction of this chapter, we recognise that the vector spaces in which X and S take values are \mathfrak{g} and \mathfrak{g}^* respectively.² Furthermore, F and H are translations acting on \mathfrak{g} and \mathfrak{g}^* respectively, and A and B are elements of the *adjoint representation* of \mathfrak{g} and its corresponding dual representation, respectively. The adjoint representation $\text{ad} : \mathfrak{g} \rightarrow \mathfrak{gl}(\mathfrak{g})$ maps Lie algebra elements to linear maps on the Lie algebra itself. We will consider X and N in their fundamental matrix representation $X, N \in \mathbb{R}^{m \times m}$ for some m , and we write their indices as X_{ij} and N_{ij} , $i, j = 1, \dots, m$.

From Eq. 7.4 we find the short-term propagators

$$X_\alpha(t_0 + \Delta t, \mathbf{u}) \approx e^{-\Delta t \text{ad}_N} X_0 + \left(\int_0^{\Delta t} e^{-(\Delta t - \Delta t') \text{ad}_N} d\Delta t' \right) F_{\alpha,0} \quad (7.11a)$$

$$S(t_0 + \Delta t, \mathbf{u}) \approx e^{-\Delta t \text{ad}_N^*} S_0 + \left(\int_0^{\Delta t} e^{-(\Delta t - \Delta t') \text{ad}_N^*} d\Delta t' \right) H_0. \quad (7.11b)$$

To use the short-term propagators the matrix exponentials and the integrals must be evaluated. Though as $\text{ad}_N \in \mathbb{R}^{m^2 \times m^2}$ in matrix form, numerically evaluating the exponentials of adjoint representations can be expensive. However, we can more efficiently compute them using the identity [266]

$$\text{Ad}_{e^Y} = e^{\text{ad}_Y} \quad (7.12)$$

for any $Y \in \mathfrak{g}$, and Ad is the adjoint action of the Lie group, defined as

$$\text{Ad}_g Y = g Y g^{-1} \quad (7.13)$$

for any $g \in G$. We therefore have that, using $(e^Y)^{-1} = e^{-Y}$

$$e^{-\Delta t \text{ad}_N} X_{\alpha,0} = e^{-\Delta t N} X_{\alpha,0} e^{\Delta t N} \quad (7.14a)$$

$$\left(\int_0^{\Delta t} e^{-(\Delta t - \Delta t') \text{ad}_N} d\Delta t' \right) F_{\alpha,0} = \int_0^{\Delta t} e^{-(\Delta t - \Delta t') N} F_{\alpha,0} e^{(\Delta t - \Delta t') N} d\Delta t' \quad (7.14b)$$

where the right-hand sides are now expressed in terms of the exponentials of the Lie algebra elements, which are computationally more convenient to compute. To find the

²That is, $V = \mathfrak{g}$ and $V = \mathfrak{g}^*$ for Eq. 7.9a and Eq. 7.9b respectively.

equivalent expressions for Eq. 7.11b, note that from $e^{\text{ad}_Y^*} = e^{(\text{ad}_Y)^T} = (e^{\text{ad}_Y})^T$ we have that $e^{\text{ad}_Y^*} = (\text{Ad}_e Y)^T$. To find the transpose of the adjoint action, we can write Eq. 7.13 in index form $(\text{Ad}_g Y)_{ij} = (\text{Ad}_g)_{ik,jl} Y_{kl}$, where $(\text{Ad}_g)_{ik,jl} = g_{ik} g_{lj}^{-1}$, summing over repeated indices. Then $(\text{Ad}_g)_{ik,jl}^T = (\text{Ad}_g)_{jl,ik} = g_{jl} g_{ki}^{-1}$ such that

$$(\text{Ad}_g)^T Z = (g^{-1})^T Z g^T \quad (7.15)$$

for any $Z \in \mathfrak{g}^*$ and $g \in G$. We thus find that

$$e^{-\Delta t \text{ ad}_N^*} S_0 = (e^{\Delta t N})^T S_0 (e^{-\Delta t N})^T \quad (7.16a)$$

$$\begin{aligned} \left(\int_0^{\Delta t} e^{-(\Delta t - \Delta t') \text{ ad}_N^*} d\Delta t' \right) H_0 &= \int_0^{\Delta t} (e^{(\Delta t - \Delta t') N})^T H_0 (e^{-(\Delta t - \Delta t') N})^T d\Delta t' \\ &= \left(\int_0^{\Delta t} (e^{-(\Delta t - \Delta t') N}) H_0^T (e^{(\Delta t - \Delta t') N}) d\Delta t' \right)^T \end{aligned} \quad (7.16b)$$

Comparing Eqs. 7.14 and 7.16, we see that they can be evaluated using the same sequence of calculations. We define

$$\mathcal{E}_G(\Delta t, N, Y) = \int_0^{\Delta t} e^{-(\Delta t - \Delta t') N} Y e^{(\Delta t - \Delta t') N} d\Delta t' \quad (7.17)$$

where $N, Y \in \mathbb{R}^{m \times m}$ are constant matrices, such that we can rewrite Eq. 7.11 as

$$X_\alpha(t_0 + \Delta t, \mathbf{u}) \approx e^{-\Delta t N} X_{\alpha,0} e^{\Delta t N} + \mathcal{E}_G(\Delta t, N, F_{\alpha,0}) \quad (7.18a)$$

$$S(t_0 + \Delta t, \mathbf{u}) \approx (e^{\Delta t N})^T S_0 (e^{-\Delta t N})^T + \mathcal{E}_G(\Delta t, N, H_0^T)^T, \quad (7.18b)$$

where the transposes in Eq. 7.18b can be understood by recalling the matrix representation of the dual Lie algebra $\mathfrak{g}^* = \{Y^T : Y \in \mathfrak{g}\}$ (see Eq. 6.11). We can thus treat Eq. 7.18b in a manner identical to Eq. 7.18a by taking the transpose, to get

$$S^T(t_0 + \Delta t, \mathbf{u}) \approx e^{-\Delta t N} S_0^T e^{\Delta t N} + \mathcal{E}_G(\Delta t, N, H_0^T). \quad (7.19)$$

We see that in order to evaluate the short-term propagator we need to compute $e^{-\Delta t N}$, $\mathcal{E}_G(\Delta t, N, F_0)$ and $\mathcal{E}_G(\Delta t, N, H_0^T)$. For many Lie groups, the exponential map is known in closed-form such that $e^{-\Delta t N}$ can be computed efficiently. If this is not the case, it can be approximated using

$$e^Y = \sum_{k=0}^{\infty} \frac{1}{k!} Y^k \quad (7.20)$$

up to arbitrary precision by truncating the sum. Once an expression for $e^{-\Delta t N}$ has been found (whether it is analytic or approximate), we must compute $\mathcal{E}_G(\Delta t, N, Y)$. In principle

this can be done using numerical quadrature, but in Sec. 7.2 and Sec. 7.3 we derive their closed-form expressions for $SO(3)$ - and $SE(3)$ -configured systems.

This concludes the derivation of the geometric integrator for general G -configured systems, as well as systems for which G is the symmetry group of their homogeneous configuration space. Equation 7.18a is in essence the result of evaluating Eq. 7.8. That is, we exponentiated the combined infinitesimal action of translations in \mathfrak{g} and the adjoint ad , to their corresponding actions defined on the Lie group. In other words, if we see the right-hand side of Eq. 7.9a as the infinitesimal action of the product Lie algebra $\mathfrak{t}(\mathfrak{g}) \times \mathfrak{g}$ in its adjoint representation,³ then the right-hand side of Eq. 7.18a is the corresponding action of $T(\mathfrak{g}) \times G$ in its adjoint representation. Using the isomorphism between \mathfrak{g} and \mathfrak{g}^* , we then constructed the corresponding geometric integrator for the latter, resulting in Eq. 7.18b. We note that the geometric integrator on the dual space could also have been derived using the identity [266]

$$\text{Ad}_{e^Y}^* = e^{-\text{ad}_Y^*} \quad (7.21)$$

for any $Y \in \mathfrak{g}$, which shows that ad_Y^* is the infinitesimal generator of the dual adjoint action of the Lie group.

7.2 Geometric integrators for $SO(3)$ -configured system

Here we construct the short-term propagator for $SO(3)$ -configured systems. The results also apply to systems with homogeneous configuration spaces, like S^2 , on which $SO(3)$ is the associated symmetry group. The closed-form expression for the exponential map $\exp_{SO(3)} : \mathfrak{so}(3) \rightarrow SO(3)$ is [267]

$$\exp_{SO(3)}(\hat{Y}) = \mathbb{1}_3 + \frac{\sin |\vec{Y}|}{|\vec{Y}|} \hat{Y} + \frac{1 - \cos |\vec{Y}|}{|\vec{Y}|^2} \hat{Y}^2 \quad (7.22)$$

where $\vec{Y} \in \mathbb{R}^3$ and $\hat{Y} \in \mathfrak{so}(3)$ is given by the hat map Eq. 4.66. Now using the identity $R\hat{Y}R^{-1} = \widehat{R\vec{Y}}$ for any $R \in SO(3)$, we have

$$\mathcal{E}_{SO(3)}(\Delta t, \hat{N}, \vec{Y}) = \Delta t \mathcal{B}(-\Delta t \hat{N}) \vec{Y} \quad (7.23)$$

³Here $\mathfrak{t}(\mathfrak{g})$ and $T(\mathfrak{g})$ refers to the Lie group of translations on the vector space \mathfrak{g} .

where

$$\begin{aligned}\mathcal{B}(\hat{Y}) &= \int_0^1 \exp_{SO(3)}(\alpha \hat{Y}) d\alpha \\ &= \mathbb{1}_3 + \frac{1 - \cos |\vec{Y}|}{|\vec{Y}|^2} \hat{Y} + \frac{|\vec{Y}| - \sin |\vec{Y}|}{|\vec{Y}|^3} \hat{Y}^2.\end{aligned}\quad (7.24)$$

For convenience we have let $\mathcal{E}_{SO(3)}(\Delta t, \hat{N}, \hat{Y}) \in \mathbb{R}^3$ using the inverse hat-map. The short-term propagator can then be written as

$$\vec{X}_\alpha(t_0 + \Delta t, \mathbf{u}) \approx \exp_{SO(3)}(-\Delta t \hat{N}) \vec{X}_{\alpha,0} + \mathcal{E}_{SO(3)}(\Delta t, \hat{N}, \vec{F}_{\alpha,0}) \quad (7.25a)$$

$$\vec{S}(t_0 + \Delta t, \mathbf{u}) \approx \exp_{SO(3)}(-\Delta t \hat{N}) \vec{S}_0 + \mathcal{E}_{SO(3)}(\Delta t, \hat{N}, \vec{H}_0). \quad (7.25b)$$

where $\vec{F}_{\alpha,0}$ and \vec{H}_0 are computed using Eq. 7.10. As the exponential map Eq. 7.24 is ill-conditioned for small $|\vec{Y}|$, it is often necessary to Taylor expand $\exp_{SO(3)}(\alpha \hat{Y})$ and $\mathcal{E}_{SO(3)}(\Delta t, \hat{N}, \vec{Y})$ to at least $O(|\vec{Y}|^6)$ when $|\vec{Y}| < 0.1$.

7.3 Geometric integrators for $SE(3)$ -configured systems

Here we construct the short-term propagator for $SE(3)$ -configured systems. The results also apply to systems with homogeneous configuration spaces, like \mathbb{E}^3 , on which $SE(3)$ is the associated symmetry group. We will only present the results here. See Appendix D for detailed derivations. As for the Cosserat systems, we write the kinematic and dynamical quantities of the equations of motion as $X_\alpha = \{\vec{\theta}_\alpha, \vec{\pi}_\alpha\}$, $N = \{\vec{V}; \vec{\Omega}\}$, $S = \{\vec{P}; \vec{L}\}^*$ and $Q = \{\vec{F}; \vec{M}\}^*$. All quantities are assumed to be expressed in a non-dimensionalised form.

The closed form expression for the exponential map $\exp_{SE(3)} : \mathfrak{se}(3) \rightarrow SE(3)$ is [267]

$$\exp_{SE(3)}(Y) = \begin{pmatrix} 1 & \vec{0}^T \\ \mathcal{B}(\hat{m})\vec{a} & \exp_{SO(3)}(\hat{m}) \end{pmatrix}. \quad (7.26)$$

In Appendix D we show that

$$\mathcal{E}_{SE(3)}(\Delta t, N, Y) = \begin{pmatrix} 0 & \vec{0}^T \\ \Delta t \mathcal{B}(-\Delta t \hat{\Omega})\vec{a} + \vec{r}(\Delta t, \vec{m}, \vec{V}) & \mathcal{E}_{SO(3)}(\Delta t, \hat{\Omega}, \hat{m}) \end{pmatrix} \quad (7.27)$$

where

$$\vec{r}(\Delta t, N, \vec{m}) = \frac{1}{|\vec{\Omega}|^2} \sum_{i,j=1}^3 R_{ij} \bar{\Omega}^{i-1} \hat{m} \bar{\Omega}^{j-1} \vec{V} \quad (7.28)$$

where $\bar{\Omega} = \hat{\Omega}/|\vec{\Omega}|$ and

$$R = \psi^{-2} \int_0^1 d\alpha P = \begin{pmatrix} \frac{\psi^2}{2} & -\psi + \sin(\psi) & \frac{\psi^2}{2} + \cos(\psi) - 1 \\ -\psi \cos(\psi) + \sin(\psi) & -\frac{(\cos(\psi)-1)^2}{2} & -\psi \cos(\psi) - \frac{\psi}{2} + \sin(\psi) + \frac{\sin(2\psi)}{4} \\ \frac{\psi^2}{2} - \psi \sin(\psi) - \cos(\psi) + 1 & -\frac{3\psi}{2} + 2 \sin(\psi) - \frac{\sin(2\psi)}{4} & \frac{(\psi-\sin(\psi))^2}{2} \end{pmatrix}, \quad (7.29)$$

where $\psi = \Delta t |\vec{\Omega}|$. Using Eq. 7.26 and Eq. 7.27, together with Eq. 7.18 the short-term propagator for $SE(3)$ -configured systems can be computed.

In practice we may want to work with the sub-matrices of the kinodynamic degrees of freedom, rather than in their full matrix form. In Appendix D we also show that the short-term propagator can be written out explicitly as

$$\vec{\theta}_\alpha(t_0 + \Delta t, \mathbf{u}) \approx U_0^1 \vec{\theta}_{0,\alpha} + \Delta t U_0^2 \vec{f}_{0,\alpha}^1 + \vec{s}_0(\vec{\pi}_{0,\alpha}) + \vec{r}(\Delta t, N_0, \vec{f}_{0,\alpha}^2) \quad (7.30a)$$

$$\vec{\pi}_\alpha(t_0 + \Delta t, \mathbf{u}) \approx U_0^1 \vec{\pi}_{0,\alpha} + \mathcal{E}_{SO(3)}(\Delta t, \hat{\Omega}, \vec{f}_{0,\alpha}^2) \quad (7.30b)$$

$$\vec{P}(t_0 + \Delta t, \mathbf{u}) \approx U_0^1 \vec{P}_0 + \Delta t U_0^2 \vec{h}_0^1 + \vec{s}_0(\vec{L}_0) + \vec{r}(\Delta t, N_0, \vec{h}_0^2) \quad (7.30c)$$

$$\vec{L}(t_0 + \Delta t, \mathbf{u}) \approx U_0^1 \vec{L}_0 + \mathcal{E}_{SO(3)}(\Delta t, \hat{\Omega}, \vec{h}_0^2). \quad (7.30d)$$

where

$$X_\alpha(t_0, \mathbf{u}) := X_{0,\alpha} = \{\vec{\theta}_{\alpha,0}; \vec{\pi}_{\alpha,0}\} \quad (7.31a)$$

$$S(t_0, \mathbf{u}) := S_0 = \{\vec{P}_0; \vec{L}_0\}^* \quad (7.31b)$$

$$F_\alpha(t_0, \mathbf{u}) := F_{0,\alpha} = \{\vec{f}_{\alpha,0}^1; \vec{f}_{\alpha,0}^2\} \quad (7.31c)$$

$$H(t_0, \mathbf{u}) := H_0 = \{\vec{h}_0^1; \vec{h}_0^2\}^* \quad (7.31d)$$

and

$$U_0^1 = \exp_{SO(3)}(-\Delta t \hat{\Omega}_0) \quad (7.32a)$$

$$U_0^2 = \mathcal{B}(-\Delta t \hat{\Omega}_0) \quad (7.32b)$$

$$\vec{s}_0(\vec{m}) = \Delta t (U_0^1 \vec{m}) \times (U_0^2 \vec{V}_0) \quad (7.32c)$$

Note that as with the $SO(3)$ -propagator, the exponential mappings involved are ill-conditioned as their arguments go to zero. Therefore, in numerical applications $\exp_{SE(3)}(Y)$ and $\mathcal{E}_{SE(3)}(\Delta t, N, Y)$ should be Taylor expanded when the arguments are small.

Finally, in Appendix D.1 we also derive an alternative short-term propagator, that exploits the fact that $\vec{\theta}_\alpha$, $\vec{\pi}_\alpha$, \vec{P} and \vec{L} can be integrated separately using the $SO(3)$ -propagator defined in Sec. 7.2. The resulting propagator enjoys a simplified form, and does not require computing $\vec{r}(\Delta t, N, \vec{m})$, which is the most costly operation required for the

full $SE(3)$ -propagator. However, the simplification comes at the cost of not encapsulating the full geometry of the equations of motion. We will compare the two version of the propagators, as well as the Forward-Euler propagator, in Sec. 7.4.

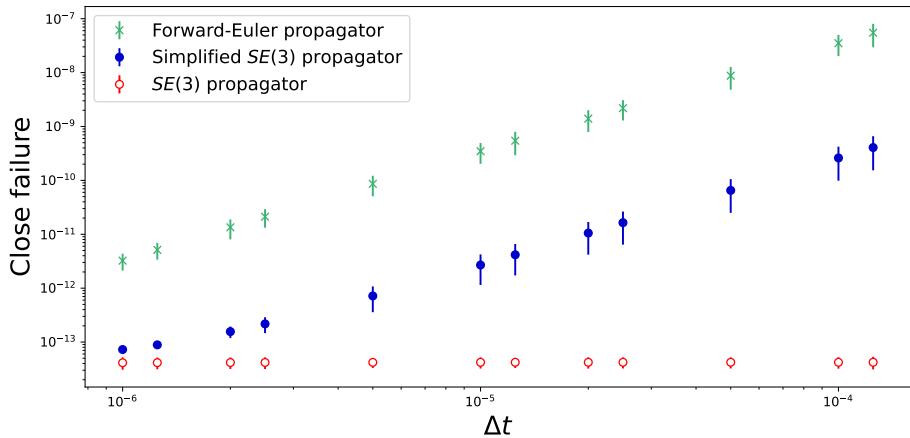
7.4 Benchmark tests

Here we compare the performance of the Forward-Euler Eq. 7.2, the $SE(3)$ and the simplified $SE(3)$ propagators in numerics. In simulations, we have observed that the latter two propagators enjoy smaller errors in general during numerical integration. However, the main benefit of geometric integrators are that they preserve the geometric structures of the systems. We consider three systems: Overdamped Cosserat rod kinodynamics (see 5.125), underdamped Cosserat rod kinodynamics (see Eq. 5.120) and Cosserat surface kinematics (see Eq. 6.92).

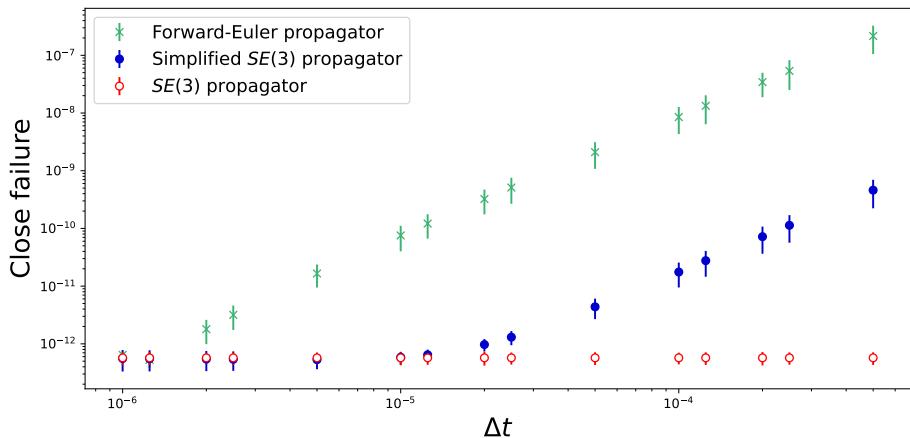
For the Cosserat rod systems, we considered a random initial *closed* configuration (see Fig. F.1 for an example), under the influence of randomly configured constitutive forces and moments. We then simulated the systems until a time T , and computed the extent to which the rod *fails to be closed* at T . See Fig. 7.2a for an illustration. In other words, if $\Phi(u, t)$ is the spatio-temporal configuration of the simulated rod, we compute the distance $|\Phi(L_0, T) - \Phi(0, T)|$, which is the *closure failure* of the rod. Due to numerical errors, rods that are initially configured as closed will in general fail to remain so. For a range of time-discretisations Δt , we simulated $N_{\text{sim}} = 50$ random configurations, and found the resulting closure failures. We then computed sample means and variances of the closure failures as a function of Δt . The results are shown in Fig. 7.1a and Fig. 7.1b for the overdamped and underdamped systems respectively. We see that the simplified $SE(3)$ propagator outperforms the Forward-Euler, however the closure error for the $SE(3)$ propagator is vanishingly small and near machine precision. We also see that the $SE(3)$ propagator is consistent in this regard, with its relatively small sample variances.

For the Cosserat surface, we considered an initial conditional of a flat sheet with a constant perpendicular material frame, that is then subjected to a random velocity and angular velocity profiles. We then computed the short-term propagator for a given time-discretisation Δt . On the resulting propagated state, we computed the supremum of the residual integrability error Δ_{uv}^{int} , given in Eq. 6.9 over the material base space. As for the Cosserat rod, we performed $N_{\text{sim}} = 50$ random simulations for each Δt , and computed the sample mean and variance of the error. The results are shown in Fig. 7.1c. Here we see again that the simplified $SE(3)$ propagator has a slightly improved performance relative to the Forward-Euler propagator. The full $SE(3)$ propagator however preserves the spatial integrability of the near machine precision for $\Delta t \leq 5 \times 10^{-5}$.

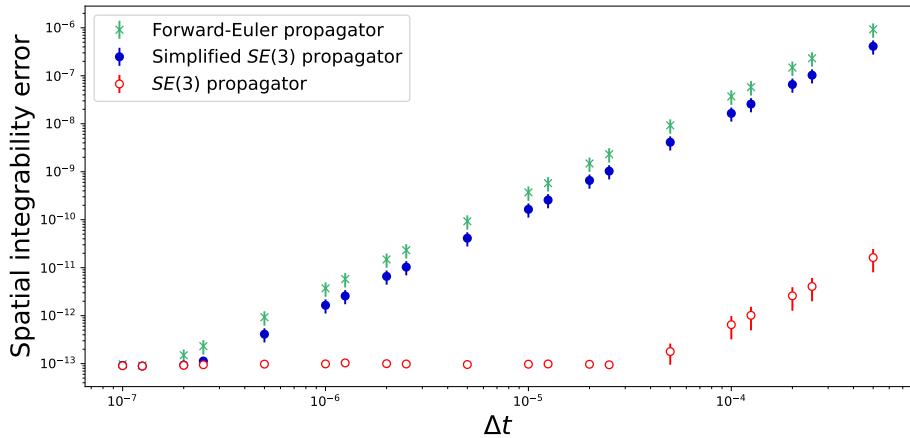
We will briefly recall the importance of the residual integrability error: If the $\Delta_{uv}^{\text{int}} \neq 0$



(a)



(b)



(c)

Figure 7.1: (a and b) The mean closure failure of a $N_{\text{sim}} = 50$ benchmark simulations of closed Cosserat rods undergoing overdamped and underdamped dynamics respectively, as a function the time-discretisation Δt , integrated using the given short-term propagators. We see that the $SE(3)$ propagator preserves the closed loop of the rod up to near-machine precision. (c) The spatial integrability error Δ_{uv}^{int} as a function Δt of a Cosserat sheet, as a result of kinematically propagating it using the given short-term propagators. We see that the $SE(3)$ propagator preserves spatial integrability where the simplified $SE(3)$ and the Forward-Euler propagators fail to do so.

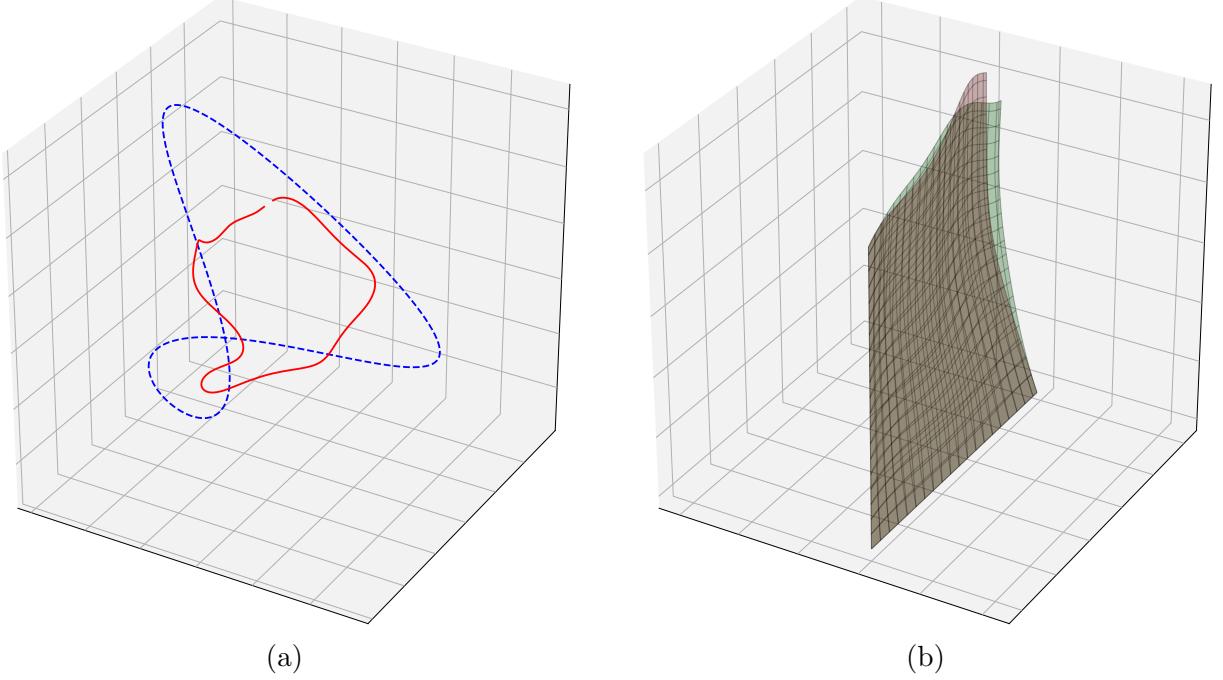


Figure 7.2: (a) Illustration of the failure of a Cosserat rod center-line to remain closed during numerical simulations. Given a closed loop initial condition $\Phi_0 = (\mathbf{r}_0; E_0)$ (dashed blue line), satisfying $\mathbf{r}_0(0) = \mathbf{r}_0(L_0)$, simulating the kinodynamic equations may lead to numerical errors that cause the final configuration to no longer be closed (solid red line). (b) Illustration of a Cosserat surface with spatial integrability fields X_u and X_v that do not obey the spatial integrability conditions Eq. 6.32b. Due to the failure of integrability, the spatio-temporal configuration $\Phi = (\mathbf{r}; E)$ is not a single-valued function of the material coordinates u and v . Instead, the spatio-temporal configuration must be considered a functional $\Phi = \Phi[\gamma_{u,v}]$, where $\gamma_{u,v} : [0, 1] \rightarrow (u, v)$ is a curve satisfying $\gamma_{u,v}(0) = (u_0, v_0)$ and $\gamma_{u,v}(1) = (u, v)$, and where $\Phi[\gamma_{u,v}]$ is the exponentiation of X_u and X_v along γ (see Appendix C for details). The full configuration is then found by computing the image of Φ under the set of all possible curves $\gamma_{u,v}$. In the figure we show the mid-surface \mathbf{r} of an example Cosserat surface that does not satisfy spatial integrability.

then the spatial reconstruction fields X_u and X_v do not satisfy the spatial integrability conditions Eq. 6.6. In practical terms, this means that X_u and X_v do not integrate to a unique spatio-temporal configuration Φ . Rather Φ is a functional $\Phi = \Phi[\gamma_{u,v}]$ of curves in the material base space $\gamma_{u,v} : [0, 1] \rightarrow (u, v)$ satisfying $\gamma_{u,v}(0) = (u_0, v_0)$ and $\gamma_{u,v}(1) = (u, v)$, along which X_u and X_v are exponentiated (see Appendix C for details). In such cases, the multi-valued configuration would be found by computing the image of Φ under the set of all possible curves $\gamma_{u,v}$. See Fig. 7.2b for an illustration. Furthermore, in Eq. 6.10 we see that integrability errors suffer exponential growth. Thus, simulations using non-geometric integrators will in general require intermittently stopping to spatially re-integrate X_α in order to maintain the errors. We thus see the importance and utility of constructing a geometric integrator for a given G -configured system.

7.5 Conclusion

The purpose of geometric integrators are to preserve qualitative features of a system, in the numerical solution of their equations of motion. For the class of systems considered in this text, it is important to preserve spatial integrability. Failure to do so causes the spatio-temporal configuration of the system to be multi-valued. Furthermore, we have previously shown that errors in spatial integrability suffer exponential growth. In this chapter we constructed geometric integrators for this purpose.

The geometric integrators we derived can be compared with the Lie-Poisson integrators of [262, 263, 264], defined on finite-dimensional systems. The primary point of difference is the presence of the translation terms F_α and H in Eq. 7.9, which we take care to exponentiate in combination with the infinitesimal action of the adjoint $-\text{ad}_N$ and its dual $-\text{ad}_N^*$ respectively. These translational terms arise due to the presence of generalised constitutive and body forces (see Eq. 7.10). Whilst a finite dimensional system can be subject to the latter, its finitude precludes constitutive dynamics (for example, recall the equations of motion of a rigid body Eq. 5.25), whilst constitutive dynamics is a fundamental feature of any real continuum system. It is therefore prudent to construct geometric integrators that accommodates both terms in Eq. 7.9, and we gave numerical evidence of the importance of doing so in Sec. 7.4. By carefully exponentiating the right-hand sides of the kinodynamic equations of motion Eq. 6.20, we found expressions for geometric integrators for general Lie group and homogeneous space-continua. We then applied our general result to consider the special cases of $SO(3)$ - and $SE(3)$ -configured continua. The resulting geometric integrators were shown in numerical experiments to preserve the spatial integrability, and we compared their performance relative to non-geometric integrators.

Our work in this chapter can be seen as a contribution to the literature on geometric

integrators, where we have applied it in a continuum setting. There is an extensive literature on which to draw on for future work [258, 259, 260, 261, 262, 263, 264]. For example, whilst the integrators we propose here successfully preserve geometric properties, errors still accrue within the sub-manifold of configurations that abide by spatial integrability. Such errors can be reduced, for example, using *Runge–Kutta–Munthe–Kaas methods* [268?], which generalise Runge-Kutta methods [269] to Lie group-configured systems.

Summary

The research presented in this thesis divides into two parts. In the first part, we studied and developed methods for studying the transition path ensemble of Itô diffusion equations. In the second part, we developed a geometric theory of continuum mechanics of systems with Lie group- or homogeneous configuration spaces. Below we summarise the main results of each chapter of the two research streams discussed in this thesis.

In Ch. 1 we began our exploration of the transition path ensemble by considering the Freidlin-Wentzell limit, which corresponds to a limit of vanishing diffusivity. We presented numerical methods for computing most probable transition paths and quasi-potentials of general Itô diffusion equations with additive noise. The method directly minimises the Freidlin-Wentzell action, which allowed for flexibility in choosing the path parametrisation. It uses numerical quadrature to reduce the action to a multivariate function, whose minimum is obtained by either gradient-free and gradient-based optimisation. This provides both the minimum action path and quasipotential. The direct method consisted of a discretisation of the Freidlin-Wentzell action, followed by a search for the minimum in the resulting finite-dimensional space. This approach is algorithmically simple and we showed its accuracy and efficiency on a number of benchmark problems.

In Ch. 2 we considered the TPE at finite temperatures. We applied recent mathematical developments in the field of functional Markov chain Monte Carlo methods to the sampling of stochastic transition pathways. On this foundation, we developed an MCMC scheme designed to sample TPEs with multiple competing transition channels, which we called the teleporter MCMC. The method was based on a semi-classical expansion of the path-probability measure around the dominant transition paths of the distribution, with which we constructed independence samplers allowing for the MCMC to intermittently ‘teleport’ between transition channels. Akin to the Ritz method of the Ch. 1, the MCMC procedures were discretised using a global spectral basis. Specifically, we expanded stochastic paths in the Kosambi-Karhunen-Loeve of the Brownian bridge process. We showed that spectrum of modes in this expansion can be separated into high- and low-frequency bands, where only the lower-band modes are non-trivially distributed, whilst the statistics of the latter is indistinguishable from that of free diffusion. Future research directions may involve devising multi-level MCMC schemes designed around band-structure of the KKL-based

methods we developed in this chapter. Furthermore, we may consider extending the algorithm to sample transition path ensembles of field theories.

In Ch. 3 we studied the concentration of competing transition channels in the transition path ensemble, as a function of diffusivity. Using two model systems, we showed that the dominant transition channel does not in general coincide with most probable paths of the path measure, even in a low-to-intermediate temperature regime. We approximated the TPE as mixed Gaussian measures using the semi-classical expansions developed in the previous chapter, using which we constructed semi-analytical approximators of channel rate probabilities. In the regime of validity of these approximators, we showed that the relative dominance of transition channels is a conspiracy between the path-probability of the instanton, reflecting the energetic cost of transition, and Gaussian normalisation constants, reflecting the fluctuations around the instanton.

In Ch. 4 we began our study of the geometric continuum mechanics of systems with Lie group- or homogeneous configuration spaces. We reviewed the classical theories that our work builds on, and developed the necessary mathematical tools for the results of the subsequent chapters. Of these, the mathematical preliminaries, though dense, by itself foreshadows and serves as a dry-run for the geometrical treatment that was to come.

In Ch. 5 we derived the kinematic and dynamic - in short, kinodynamic - equations of Cosserat rods and filaments. These were chosen as paradigmatic examples of systems with Lie group- and homogeneous configuration spaces, and served as a foundation for the generalisations of the chapter that followed. We identified the configuration space of the Cosserat rod as the special Euclidean group $SE(3)$, and used the Euler-Poincaré theorem to find conservative force and moment balance equations. Furthermore, we constructed a generalised Lagrange-D'Alembert principle from which arbitrary non-conservative dynamics can be derived, and expressed the resulting kinodynamic equations of motion in terms of a spatial reconstruction field X and generalised momentum field S , which are defined on the Lie algebra and its dual respectively. This Lie algebraic formulation was shown to naturally lead to kinodynamics expressed in terms of the intrinsic geometry of the rod. We then studied the filament model, which we devised as a Cosserat rod subject to kinematic constraints, and found its kinodynamic equations of motion. We thus demonstrated how systems with homogeneous configuration spaces can be obtained by imposing kinematic constraints on Lie group-configured systems.

In Ch. 6 we studied generalised Cosserat systems, which we defined as continuum bodies with a material base of general topology, and Lie group- or homogeneous configuration spaces. We formulated a general theory of the kinodynamics of such systems, which we called a generalised geometric Cosserat theory (GGCT). The cornerstone of the programme was to use the exponential map to relate the spatio-temporal configuraiton of the system to Lie algebra-valued reconstruction fields. This allowed us to express the kinodynamics

of generalised Cosserat systems directly in terms of its intrinsic geometry. We applied the GGCT to derive the kinodynamic equations of motion for a variety of known and new systems, which included the classical suite of Cosserat surfaces and bodies, Cosserat rods on spheres and in Minkowski space, as well as an application for non-linear σ field theory.

In Ch. 7 we developed geometric integrators for generalised Cosserat systems. We saw this as an application of Lie group integration theory to the infinite-dimensional setting of continuum systems. We demonstrated that our integrators preserve spatial integrability using the example of a Cosserat surface, and compared their performance to standard non-geometrical numeric integrators. We also showed that other qualitative features, like the closure of a rod, are preserved by our integrators.

References

- [1] Eugene Maurice Pierre Cosserat and F Cosserat. *Theory of Deformable Bodies*. National Aeronautics and Space Administration, 1909.
- [2] Alexander D Wentzell and Mark Iosifovich Freidlin. On small random perturbations of dynamical systems. *Russian Mathematical Surveys*, 25(1):1–55, February 1970. doi: 10.1070/rm1970v025n01abeh001254.
- [3] R. Graham. Statistical theory of instabilities in stationary nonequilibrium systems with applications to lasers and nonlinear optics. In G. Höhler, editor, *Springer Tracts in Modern Physics: Ergebnisse Der Exakten Naturwissenschaften; Volume 66*, pages 1–97. Springer Berlin Heidelberg, Berlin, Heidelberg, 1973. ISBN 978-3-662-40468-3. doi: 10.1007/978-3-662-40468-3_1.
- [4] R. Graham. Macroscopic potentials, bifurcations and noise in dissipative systems. In L. Garrido, editor, *Fluctuations and Stochastic Phenomena in Condensed Matter*, pages 1–34. Springer Berlin Heidelberg, Berlin, Heidelberg, 1987. ISBN 978-3-540-47401-2.
- [5] Peter G. Bolhuis, David Chandler, Christoph Dellago, and Phillip L. Geissler. Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annual Review of Physical Chemistry*, 53(1):291–318, October 2002. ISSN 0066-426X, 1545-1593. doi: 10.1146/annurev.physchem.53.082301.113146.
- [6] Rosalind J Allen, Patrick B Warren, and Pieter Rein Ten Wolde. Sampling rare switching events in biochemical networks. *Physical Review Letters*, 94(1):018104, 2005.
- [7] Rosalind J Allen, Chantal Valeriani, and Pieter Rein ten Wolde. Forward flux sampling for rare event simulations. *Journal of Physics: Condensed Matter*, 21(46):463102, 2009.
- [8] Lasse Ebener, Georgios Margazoglou, Jan Friedrich, Luca Biferale, and Rainer Grauer. Instanton based importance sampling for rare events in stochastic PDEs.

Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(6):063102, June 2019.
ISSN 1054-1500, 1089-7682. doi: 10.1063/1.5085119.

- [9] Liam Paninski. The most likely voltage path and large deviations approximations for integrate-and-fire neurons. *Journal of Computational Neuroscience*, 21(1):71–87, August 2006. ISSN 0929-5313. doi: 10.1007/s10827-006-7200-4.
- [10] Sui Huang. The molecular and mathematical basis of Waddington’s epigenetic landscape: A framework for post-Darwinian biology? *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 34(2):149–157, February 2012. ISSN 1521-1878. doi: 10.1002/bies.201100031.
- [11] Freddy Bouchet and Julien Reygner. Generalisation of the Eyring–Kramers transition rate formula to irreversible diffusion processes. *Annales Henri Poincaré*, 17(12):3499–3532, December 2016. ISSN 1424-0661. doi: 10.1007/s00023-016-0507-4.
- [12] Robert S Maier and Daniel L Stein. A scaling theory of bifurcations in the symmetric weak-noise escape problem. *Journal of Statistical Physics*, 83(3-4):291–357, 1996.
- [13] Peter G. Wolynes, Jose N. Onuchic, and D. Thirumalai. Navigating the Folding Routes. *Science*, March 1995. doi: 10.1126/science.7886447.
- [14] Ben C. Nolting and Karen C. Abbott. Balls, cups, and quasi-potentials: Quantifying stability in stochastic systems. *Ecology*, 97(4):850–864, 2016. ISSN 1939-9170. doi: 10.1890/15-1047.1.
- [15] M. Mangel. Barrier Transitions Driven by Fluctuations, with Applications to Ecology and Evolution. *Theoretical Population Biology*, 45(1):16–40, February 1994. ISSN 0040-5809. doi: 10.1006/tpbi.1994.1002.
- [16] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342, January 2000. ISSN 1476-4687. doi: 10.1038/35002131.
- [17] C.L. DeMarco. A phase transition model for cascading network failure. *IEEE Control Systems Magazine*, 21(6):40–51, December 2001. ISSN 1941-000X. doi: 10.1109/37.969134.
- [18] Randolph Nelson. Stochastic catastrophe theory in computer performance modeling. *Journal of the ACM (JACM)*, 34(3):661–685, 1987.
- [19] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, August 2011. ISBN 978-0-08-047536-3.

- [20] Crispin Gardiner. *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer Berlin Heidelberg, October 2010. ISBN 978-3-642-08962-6.
- [21] Hannes Risken and Till Frank. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer Science & Business Media, December 2012. ISBN 978-3-642-61544-3.
- [22] A. T. Bharucha-Reid. *Elements of the Theory of Markov Processes and Their Applications*. Courier Corporation, April 2012. ISBN 978-0-486-15035-2.
- [23] P. Faccioli, M. Sega, F. Pederiva, and H. Orland. Dominant Pathways in Protein Folding. *Physical Review Letters*, 97(10):108101, September 2006. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.97.108101.
- [24] Juyong Lee, In-Ho Lee, InSuk Joung, Jooyoung Lee, and Bernard R. Brooks. Finding multiple reaction pathways via global optimization of action. *Nature Communications*, 8(1):15443, August 2017. ISSN 2041-1723. doi: 10.1038/ncomms15443.
- [25] L. Onsager and S. Machlup. Fluctuations and Irreversible Processes. *Physical Review*, 91(6):1505–1512, September 1953. ISSN 0031-899X. doi: 10.1103/PhysRev.91.1505.
- [26] A. Bach, D. Dürr, and B. Stawicki. Functionals of paths of a diffusion process and the Onsager-Machlup function. *Zeitschrift für Physik B Condensed Matter and Quanta*, 26(2):191–193, June 1977. ISSN 0340-224X, 1434-6036. doi: 10.1007/BF01325272.
- [27] Hidemi Ito. Probabilistic Construction of Lagrangean of Diffusion Process and Its Application. *Progress of Theoretical Physics*, 59(3):725–741, March 1978. ISSN 0033-068X. doi: 10.1143/PTP.59.725.
- [28] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. Elsevier, June 2014. ISBN 978-1-4832-9615-9.
- [29] R. L. Stratonovich. Some Markov methods in the theory of stochastic processes in nonlinear dynamical systems. In Frank Moss and P. V. E. McClintock, editors, *Noise in Nonlinear Dynamical Systems: Volume 1: Theory of Continuous Fokker-Planck Systems*, volume 1, pages 16–71. Cambridge University Press, Cambridge, 1989. ISBN 978-0-521-11850-7. doi: 10.1017/CBO9780511897818.004.
- [30] Robert Graham. Macroscopic potentials, bifurcations and noise in dissipative systems. In Frank Moss and P. V. E. McClintock, editors, *Noise in Nonlinear Dynamical Systems: Volume 1: Theory of Continuous Fokker-Planck Systems*, volume 1, pages 225–278. Cambridge University Press, Cambridge, 1989. ISBN 978-0-521-11850-7. doi: 10.1017/CBO9780511897818.009.

- [31] Ludwig Arnold. *Stochastic Differential Equations: Theory and Applications*. Wiley, April 1974. ISBN 978-0-471-03359-2.
- [32] Jannes Gladrow, Ulrich F. Keyser, R. Adhikari, and Julian Kappler. Experimental Measurement of Relative Path Probabilities and Stochastic Actions. *Physical Review X*, 11(3):031022, July 2021. doi: 10.1103/PhysRevX.11.031022.
- [33] I. M. Gel'fand and A. M. Yaglom. Integration in Functional Spaces and its Applications in Quantum Physics. *Journal of Mathematical Physics*, 1(1):48–69, January 1960. ISSN 0022-2488. doi: 10.1063/1.1703636.
- [34] Daniel Nickelsen and Hugo Touchette. Noise correction of large deviations with anomalous scaling. *Physical Review E*, 105(6):064102, June 2022. doi: 10.1103/PhysRevE.105.064102.
- [35] Giulio Corazza and Matteo Fadel. Normalized Gaussian path integrals. *Physical Review E*, 102(2):022135, August 2020. doi: 10.1103/PhysRevE.102.022135.
- [36] Yulong Lu, Andrew Stuart, and Hendrik Weber. Gaussian Approximations for Transition Paths in Brownian Dynamics. *SIAM Journal on Mathematical Analysis*, 49(4):3005–3047, January 2017. ISSN 0036-1410. doi: 10.1137/16M1071845.
- [37] Lukas Kikuchi, Ronojoy Adhikari, and Julian Kappler. Diffusivity dependence of the transition path ensemble, August 2022.
- [38] Hongri Gu, Quentin Boehler, Haoyang Cui, Eleonora Secchi, Giovanni Savorana, Carmela De Marco, Simone Gervasoni, Quentin Peyron, Tian-Yun Huang, Salvador Pane, Ann M. Hirt, Daniel Ahmed, and Bradley J. Nelson. Magnetic cilia carpets with programmable metachronal waves. *Nature Communications*, 11(1):2637, May 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-16458-4.
- [39] Shankar Krishnaswamy. A cosserat-type model for the red blood cell wall. *International Journal of Engineering Science*, 34(8):873–899, June 1996. ISSN 0020-7225. doi: 10.1016/0020-7225(95)00139-5.
- [40] Giulio Corazza and Raushan Singh. Unraveling looping efficiency of stochastic Cosserat polymers. *Physical Review Research*, 4(1):013071, January 2022. doi: 10.1103/PhysRevResearch.4.013071.
- [41] A. Rajan, R. Pramanik, A. Narayanan, and A. Arockiarajan. Mechanics of viscoelastic buckling in slender hydrogels. *Materials Research Express*, 6(5):055320, February 2019. ISSN 2053-1591. doi: 10.1088/2053-1591/ab0691.

- [42] Jerrold E. Marsden and Tudor S. Ratiu. *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*. Springer Science & Business Media, March 2013. ISBN 978-0-387-21792-5.
- [43] E Weinan, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. *Physical Review B*, 66(5):052301, 2002.
- [44] Eric Vanden-Eijnden and Matthias Heymann. The geometric minimum action method for computing minimum energy paths. *The Journal of Chemical Physics*, 128(6):061103, February 2008. ISSN 0021-9606. doi: 10.1063/1.2833040.
- [45] Tobias Grafke, Tobias Schäfer, and Eric Vanden-Eijnden. Long Term Effects of Small Random Perturbations on Dynamical Systems: Theoretical and Computational Tools. In Roderick Melnik, Roman Makarov, and Jacques Belair, editors, *Recent Progress and Modern Challenges in Applied Mathematics, Modeling and Computational Science*, Fields Institute Communications, pages 17–55. Springer, New York, NY, 2017. ISBN 978-1-4939-6969-2. doi: 10.1007/978-1-4939-6969-2_2.
- [46] I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Courier Corporation, April 2012. ISBN 978-0-486-13501-4.
- [47] L.V. Kantorovich and V.I. Krylov. *Approximate Methods of Higher Analysis*. P. Noordhoff, Groningen, 1958.
- [48] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Minimum action method for the study of rare events. *Communications on Pure and Applied Mathematics*, 57(5):637–656, May 2004. ISSN 0010-3640. doi: 10.1002/cpa.20005.
- [49] Walter Ritz. über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik. *Journal für Mathematik*, 135:1–61, 1909.
- [50] MK Cameron. Finding the quasipotential for nongradient SDEs. *Physica D: Nonlinear Phenomena*, 241(18):1532–1550, 2012.
- [51] Shuo Yang, Samuel F Potter, and Maria K Cameron. Computing the quasipotential for nongradient SDEs in 3D. *Journal of Computational Physics*, 379:325–350, 2019.
- [52] Daisy Dahiya and Maria Cameron. Ordered Line Integral Methods for Computing the Quasi-Potential. *Journal of Scientific Computing*, 75(3):1351–1384, June 2018. ISSN 1573-7691. doi: 10.1007/s10915-017-0590-9.
- [53] Lukas Takaaki Lukas Kikuchi and Rajesh Singh. PyRitz, November 2022.

- [54] Lukas Kikuchi, Rajesh Singh, M. E. Cates, and R. Adhikari. Ritz method for transition paths and quasipotentials of rare diffusive events. *Physical Review Research*, 2(3):033208, August 2020. doi: 10.1103/PhysRevResearch.2.033208.
- [55] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer Science & Business Media, 2003. ISBN 978-3-540-04758-2.
- [56] Steven Shreve. *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*. Springer Science & Business Media, June 2005. ISBN 978-0-387-24968-1.
- [57] Grigorios A. Pavliotis. *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. Springer New York, November 2014. ISBN 978-1-4939-1322-0.
- [58] Hugo Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1):1–69, July 2009. ISSN 0370-1573. doi: 10.1016/j.physrep.2009.05.002.
- [59] Donald Ludwig. Persistence of dynamical systems under random perturbations. *Siam Review*, 17(4):605–640, 1975.
- [60] E. T. Whittaker. *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1988. doi: 10.1017/CBO9780511608797.
- [61] L. D. Landau and E. M. Lifshitz. *Mechanics (Third Edition)*. Butterworth-Heinemann, Oxford, 1976. ISBN 978-0-7506-2896-9.
- [62] Feliks Gantmacher. *Lectures in Analytical Mechanics.(Translated from the Russian by G. Yankovsky)*. Moscow: Mir Publishers, 1970.
- [63] Roberto Olander and Ron Elber. Yet another look at the steepest descent path. *Journal of Molecular Structure: THEOCHEM*, 398:63–71, 1997.
- [64] Matthias Heymann and Eric Vanden-Eijnden. The geometric minimum action method: A least action principle on the space of curves. *Communications on Pure and Applied Mathematics*, 61(8):1052–1117, 2008. ISSN 1097-0312. doi: 10.1002/cpa.20238.
- [65] Lev Ernestovich Elsgolts and George Yankovsky. *Differential Equations and the Calculus of Variations*. Mir Moscow, 1973.
- [66] Martin J Gander and Gerhard Wanner. From euler, ritz, and galerkin to modern computing. *Siam Review*, 54(4):627–666, 2012.
- [67] Lloyd N Trefethen. *Spectral Methods in MATLAB*, volume 10. Siam, 2000.

- [68] John P Boyd. *Chebyshev and Fourier Spectral Methods*. Courier Corporation, Mineola, 2001.
- [69] Lloyd N Trefethen. *Approximation Theory and Approximation Practice*, volume 128. Siam, 2013.
- [70] Richard Hamming. *Numerical Methods for Scientists and Engineers*. Courier Corporation, New York, 2012.
- [71] Herbert E Salzer. Lagrangian interpolation at the chebyshev points x_n , $\nu \equiv \cos(\nu\pi/n)$, $\nu = 0 \dots n$; some unnoted advantages. *The Computer Journal*, 15(2):156–159, 1972.
- [72] Jean-Paul Berrut and Lloyd N. Trefethen. Barycentric Lagrange Interpolation. *SIAM Review*, 46(3):501–517, January 2004. ISSN 0036-1445, 1095-7200. doi: 10.1137/S0036144502417715.
- [73] Jared L Aurentz and Lloyd N Trefethen. Chopping a chebyshev series. *ACM Transactions on Mathematical Software (TOMS)*, 43(4):33, 2017.
- [74] M. J. D. Powell. The NEWUOA software for unconstrained optimization without derivatives. In G. Di Pillo and M. Roma, editors, *Large-Scale Nonlinear Optimization*, pages 255–297. Springer US, Boston, MA, 2006. ISBN 978-0-387-30065-8. doi: 10.1007/0-387-30065-1_16.
- [75] Dieter Kraft. Algorithm 733: TOMP-Fortran modules for optimal control calculations. *ACM Transactions on Mathematical Software*, 20(3):262–281, 1994.
- [76] Steven G Johnson. The NLopt nonlinear-optimization package, 2014.
- [77] Klaus Müller and Leo D Brown. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theoretica chimica acta*, 53(1):75–93, 1979.
- [78] David Marin Roma, Ruadhan A O’Flanagan, Andrei E Ruckenstein, Anirvan M Sengupta, and Ranjan Mukhopadhyay. Optimal path to epigenetic switching. *Physical Review E*, 71(1):011902, 2005.
- [79] J Egger. Stochastically driven large-scale circulations with multiple equilibria. *Journal of the Atmospheric Sciences*, 38(12):2606–2618, 1981.
- [80] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer Science & Business Media, June 2011. ISBN 978-3-540-54062-5.

- [81] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. *Statistical Science*, 28(3):424–446, August 2013. ISSN 0883-4237, 2168-8745. doi: 10.1214/13-STS421.
- [82] Alexandros Beskos, Gareth Roberts, Andrew Stuart, and Jochen Voss. MCMC METHODS FOR DIFFUSION BRIDGES. *Stochastics and Dynamics*, 08(03):319–350, September 2008. ISSN 0219-4937, 1793-6799. doi: 10.1142/S0219493708002378.
- [83] M. Hairer, A. M. Stuart, J. Voss, and P. Wiberg. Analysis of SPDEs arising in path sampling. Part I: The Gaussian case. *Communications in Mathematical Sciences*, 3(4):587–603, December 2005. ISSN 1539-6746, 1945-0796.
- [84] M. Hairer, A. M. Stuart, and J. Voss. Analysis of SPDEs arising in path sampling part II: The nonlinear case. *The Annals of Applied Probability*, 17(5-6):1657–1706, October 2007. ISSN 1050-5164, 2168-8737. doi: 10.1214/07-AAP441.
- [85] Martin Hairer, Andrew M. Stuart, and Sebastian J. Vollmer. Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *The Annals of Applied Probability*, 24(6):2455–2490, December 2014. ISSN 1050-5164, 2168-8737. doi: 10.1214/13-AAP982.
- [86] D. D Kosambi. Parallelism and path-spaces. In *Ramaswamy R. (Eds) D.D. Kosambi*, pages 59–70. Springer, New Delhi, 2016. ISBN 978-81-322-3674-0.
- [87] Kari Karhunen. *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*. PhD thesis, Universitat Helsinki, Helsinki, 1947.
- [88] M. Loève. *Probability Theory I*, volume 45 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, fourth edition, 1977. ISBN 978-0-387-90210-4. doi: 10.1007/978-1-4684-9464-8.
- [89] M. Chaichian and A. Demichev. *Path Integrals in Physics: Volume I Stochastic Processes and Quantum Mechanics*. CRC Press, July 2001. ISBN 978-0-7503-0801-4.
- [90] L. S. Schulman. *Techniques and Applications of Path Integration*. Wiley, August 1996. ISBN 978-0-471-16610-8.
- [91] V. V. Smirnov. On the estimation of a path integral by means of the saddle point method. *Journal of Physics A: Mathematical and Theoretical*, 43(46):465303, October 2010. ISSN 1751-8121. doi: 10.1088/1751-8113/43/46/465303.
- [92] Cécile Morette. On the Definition and Approximation of Feynman’s Path Integrals. *Physical Review*, 81(5):848–852, March 1951. ISSN 0031-899X. doi: 10.1103/PhysRev.81.848.

- [93] M. S. Marinov. Path integrals in quantum theory: An outlook of basic concepts. *Physics Reports*, 60(1):1–57, April 1980. ISSN 0370-1573. doi: 10.1016/0370-1573(80)90111-8.
- [94] J. J. Sakurai and Jim Napolitano. *Modern Quantum Mechanics*. Cambridge University Press, September 2017. ISBN 978-1-108-42241-3.
- [95] Cristian Sminchisescu, Max Welling, and Geoffrey Hinton. A Mode-Hopping MCMC sampler. page 18.
- [96] Michael Lindsey, Jonathan Weare, and Anna Zhang. Ensemble Markov chain Monte Carlo with teleporting walkers, June 2021.
- [97] Christoph Dellago, Peter G. Bolhuis, Félix S. Csajka, and David Chandler. Transition path sampling and the calculation of rate constants. *The Journal of Chemical Physics*, 108(5):1964–1977, February 1998. ISSN 0021-9606. doi: 10.1063/1.475562.
- [98] Christoph Dellago, Peter G. Bolhuis, and David Chandler. On the calculation of reaction rate constants in the transition path ensemble. *The Journal of Chemical Physics*, 110(14):6617–6625, April 1999. ISSN 0021-9606. doi: 10.1063/1.478569.
- [99] Christoph Dellago, Peter G. Bolhuis, and David Chandler. Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements. *The Journal of Chemical Physics*, 108(22):9236–9245, June 1998. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.476378.
- [100] Peter G. Bolhuis and David W. H. Swenson. Transition Path Sampling as Markov Chain Monte Carlo of Trajectories: Recent Algorithms, Software, Applications, and Future Outlook. *Advanced Theory and Simulations*, 4:2000237. ISSN 2513-0390. doi: 10.1002/adts.202000237.
- [101] Fernando A. Escobedo, Ernesto E. Borrero, and Juan C. Araque. Transition path sampling and forward flux sampling. Applications to biological systems. *Journal of Physics: Condensed Matter*, 21(33):333101, July 2009. ISSN 0953-8984. doi: 10.1088/0953-8984/21/33/333101.
- [102] Rosalind J. Allen, Daan Frenkel, and Pieter Rein ten Wolde. Forward flux sampling-type schemes for simulating rare events: Efficiency analysis. *The Journal of Chemical Physics*, 124(19):194111, May 2006. ISSN 0021-9606. doi: 10.1063/1.2198827.
- [103] Sarwar Hussain and Amir Haji-Akbari. Studying rare events using forward-flux sampling: Recent breakthroughs and future outlook. *The Journal of Chemical Physics*, 152(6):060901, February 2020. ISSN 0021-9606. doi: 10.1063/1.5127780.

- [104] David J. Earl and Michael W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916, November 2005. ISSN 1463-9084. doi: 10.1039/B509983H.
- [105] Hiroshi Fujisaki, Motoyuki Shiga, and Akinori Kidera. Onsager–Machlup action-based path sampling and its combination with replica exchange for diffusive and multiple pathways. *The Journal of Chemical Physics*, 132(13):134101, April 2010. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.3372802.
- [106] Lars Holden. Mixing of MCMC algorithms. *Journal of Statistical Computation and Simulation*, 89(12):2261–2279, August 2019. ISSN 0094-9655. doi: 10.1080/00949655.2019.1615064.
- [107] Frank Moss and P. V. E. McClintock. *Noise in Nonlinear Dynamical Systems: Volume 1, Theory of Continuous Fokker-Planck Systems*. Cambridge University Press, April 1989. ISBN 978-0-521-35228-4.
- [108] Katharine L. C. Hunt and John Ross. Path integral solutions of stochastic equations for nonlinear irreversible processes: The uniqueness of the thermodynamic Lagrangian. *The Journal of Chemical Physics*, 75(2):976–984, July 1981. ISSN 0021-9606. doi: 10.1063/1.442098.
- [109] Artur B. Adib. Stochastic actions for diffusive dynamics: Reweighting, sampling, and minimization. *The Journal of Physical Chemistry B*, 112(19):5910–5916, May 2008. ISSN 1520-6106, 1520-5207. doi: 10.1021/jp0751458.
- [110] Tobias Grafke and Eric Vanden-Eijnden. Numerical computation of rare events via large deviation theory. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(6):063118, June 2019. ISSN 1054-1500, 1089-7682. doi: 10.1063/1.5084025.
- [111] Detlef Dürr and Alexander Bach. The Onsager-Machlup function as Lagrangian for the most probable path of a diffusion process. *Communications in Mathematical Physics*, 60(2):153–170, June 1978. ISSN 1432-0916. doi: 10.1007/BF01609446.
- [112] Ruslan Leontievich Stratonovich. On the probability functional of diffusion processes. *Selected Trans. in Math. Stat. Prob*, 10:273–286, 1971.
- [113] Takahiko Fujita and Shin-ichi Kotani. The Onsager-Machlup function for diffusion processes. *Journal of Mathematics of Kyoto University*, 22(1):115–130, January 1982. ISSN 0023-608X. doi: 10.1215/kjm/1250521863.
- [114] W. Horsthemke and A. Bach. Onsager-Machlup Function for one dimensional nonlinear diffusion processes. *Zeitschrift für Physik B Condensed Matter*, 22(2):189–192, June 1975. ISSN 1431-584X. doi: 10.1007/BF01322364.

- [115] *Measure and Integration Theory on Infinite-Dimensional Spaces: Abstract Harmonic Analysis*. Academic Press, October 1972. ISBN 978-0-08-087363-3.
- [116] Leonard Gross. Abstract Wiener spaces. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory, Part 1*, volume 5.2A of *Berkeley Symposium on Mathematical Statistics and Probability*, pages 31–42. University of California Press, California, 1967.
- [117] R. H. Cameron and W. T. Martin. Transformations of Wiener Integrals Under Translations. *Annals of Mathematics*, 45(2):386–396, 1944. ISSN 0003-486X. doi: 10.2307/1969276.
- [118] R. H. Cameron and W. T. Martin. Transformations of Wiener Integrals Under a General Class of Linear Transformations. *Transactions of the American Mathematical Society*, 58(2):184–219, 1945. ISSN 0002-9947. doi: 10.2307/1990282.
- [119] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, June 1953. ISSN 0021-9606. doi: 10.1063/1.1699114.
- [120] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, April 1970. ISSN 0006-3444. doi: 10.1093/biomet/57.1.97.
- [121] Richard Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961. ISBN 978-0-691-07901-1.
- [122] J. Makhoul. A fast cosine transform in one and two dimensions. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(1):27–34, February 1980. ISSN 0096-3518. doi: 10.1109/TASSP.1980.1163351.
- [123] N. Ahmed, T. Natarajan, and K.R. Rao. Discrete Cosine Transform. *IEEE Transactions on Computers*, C-23(1):90–93, January 1974. ISSN 1557-9956. doi: 10.1109/T-C.1974.223784.
- [124] Eugene Wong and Moshe Zakai. On the Convergence of Ordinary Integrals to Stochastic Integrals. *The Annals of Mathematical Statistics*, 36(5):1560–1564, October 1965. ISSN 0003-4851, 2168-8990. doi: 10.1214/aoms/1177699916.
- [125] Krystyna Twardowska. Wong-Zakai approximations for stochastic differential equations. *Acta Applicandae Mathematica*, 43(3):317–359, June 1996. ISSN 1572-9036. doi: 10.1007/BF00047670.

- [126] Peter K. Friz and Nicolas B. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-87607-0. doi: 10.1017/CBO9780511845079.
- [127] A. Sokal. Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms. In Cecile DeWitt-Morette, Pierre Cartier, and Antoine Folacci, editors, *Functional Integration: Basics and Applications*, NATO ASI Series, pages 131–192. Springer US, Boston, MA, 1997. ISBN 978-1-4899-0319-8. doi: 10.1007/978-1-4899-0319-8_6.
- [128] Jonathan Goodman and Jonathan Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5:65–80, January 2010. doi: 10.2140/camcos.2010.5.65.
- [129] Michael B. Giles. Multilevel Monte Carlo Path Simulation. *Operations Research*, 56(3):607–617, June 2008. ISSN 0030-364X. doi: 10.1287/opre.1070.0496.
- [130] T. J. Dodwell, C. Ketelsen, R. Scheichl, and A. L. Teckentrup. A Hierarchical Multilevel Markov Chain Monte Carlo Algorithm with Applications to Uncertainty Quantification in Subsurface Flow. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):1075–1108, January 2015. doi: 10.1137/130915005.
- [131] Karl Jansen, Eike H. Müller, and Robert Scheichl. Multilevel Monte Carlo algorithm for quantum mechanics on a lattice. *Physical Review D*, 102(11):114512, December 2020. ISSN 2470-0010, 2470-0029. doi: 10.1103/PhysRevD.102.114512.
- [132] Paul B. Rohrbach, Hideki Kobayashi, Robert Scheichl, Nigel B. Wilding, and Robert L. Jack. Multilevel simulation of hard-sphere mixtures. *The Journal of Chemical Physics*, 157(12):124109, September 2022. ISSN 0021-9606. doi: 10.1063/5.0102875.
- [133] Michael E. Cates and Elsen Tjhung. Theories of binary fluid mixtures: From phase-separation kinetics to active emulsions. *Journal of Fluid Mechanics*, 836:P1, February 2018. ISSN 0022-1120, 1469-7645. doi: 10.1017/jfm.2017.832.
- [134] Adriano Tiribocchi, Raphael Wittkowski, Davide Marenduzzo, and Michael E. Cates. Active Model H: Scalar Active Matter in a Momentum-Conserving Fluid. *Physical Review Letters*, 115(18):188302, October 2015. doi: 10.1103/PhysRevLett.115.188302.
- [135] Joakim Stenhammar, Adriano Tiribocchi, Rosalind J. Allen, Davide Marenduzzo, and Michael E. Cates. Continuum Theory of Phase Separation Kinetics for Active

- Brownian Particles. *Physical Review Letters*, 111(14):145702, October 2013. doi: 10.1103/PhysRevLett.111.145702.
- [136] Raphael Wittkowski, Adriano Tiribocchi, Joakim Stenhammar, Rosalind J. Allen, Davide Marenduzzo, and Michael E. Cates. Scalar Φ^4 field theory for active-particle phase separation. *Nature Communications*, 5(1):4351, July 2014. ISSN 2041-1723. doi: 10.1038/ncomms5351.
- [137] David S. Dean. Langevin equation for the density of a system of interacting Langevin processes. *Journal of Physics A: Mathematical and General*, 29(24):L613, December 1996. ISSN 0305-4470. doi: 10.1088/0305-4470/29/24/001.
- [138] Michael E. Cates. Active Field Theories, April 2019.
- [139] Crispin Gardiner. *Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences*. Springer Berlin Heidelberg, 1990. ISBN 978-3-540-15607-9.
- [140] Limin Wang. *Karhunen-Loeve Expansions and Their Applications*. London School of Economics and Political Science (United Kingdom), 2008.
- [141] Sylvain Corlay. Properties of the Ornstein-Uhlenbeck bridge. *arXiv:1310.5617 [math]*, January 2014.
- [142] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 1995.
- [143] David W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, March 2015. ISBN 978-0-471-69755-8.
- [144] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, November 2016. ISBN 978-0-262-33737-3.
- [145] T. Tin Nguyen, Hien D. Nguyen, Faicel Chamroukhi, and Geoffrey J. McLachlan. Approximation by finite mixtures of continuous density functions that vanish at infinity. *Cogent Mathematics & Statistics*, 7(1):1750861, January 2020. ISSN null. doi: 10.1080/25742558.2020.1750861.
- [146] M.A. Carreira-Perpinan. Mode-finding for mixtures of Gaussian distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1318–1323, November 2000. ISSN 1939-3539. doi: 10.1109/34.888716.
- [147] Tobias Grafke, Rainer Grauer, and Tobias Schäfer. The instanton method and its numerical implementation in fluid mechanics. *Journal of Physics A: Mathematical and Theoretical*, 48(33):333001, August 2015. ISSN 1751-8113, 1751-8121. doi: 10.1088/1751-8113/48/33/333001.

- [148] Timo Schorlepp, Tobias Grafke, and Rainer Grauer. Gel'fand–Yaglom type equations for calculating fluctuations around instantons in stochastic systems. *Journal of Physics A: Mathematical and Theoretical*, 54(23):235003, June 2021. ISSN 1751-8113, 1751-8121. doi: 10.1088/1751-8121/abfb26.
- [149] Giovanni Dematteis, Tobias Grafke, Miguel Onorato, and Eric Vanden-Eijnden. Experimental Evidence of Hydrodynamic Instantons: The Universal Route to Rogue Waves. *Physical Review X*, 9(4):041057, December 2019. ISSN 2160-3308. doi: 10.1103/PhysRevX.9.041057.
- [150] Grégoire Ferré and Tobias Grafke. Approximate optimal controls via instanton expansion for low temperature free energy computation. *arXiv:2011.10990 [cond-mat]*, May 2021.
- [151] James F. Lutsko. How crystals form: A theory of nucleation pathways. *Science Advances*, April 2019. doi: 10.1126/sciadv.aav7399.
- [152] Pieter Rein ten Wolde, Maria J. Ruiz-Montero, and Daan Frenkel. Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling. *The Journal of Chemical Physics*, 104(24):9932–9947, June 1996. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.471721.
- [153] Weiqing Ren, Eric Vanden-Eijnden, Paul Maragakis, and Weinan E. Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide. *The Journal of Chemical Physics*, 123(13):134109, October 2005. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.2013256.
- [154] Hiroshi Fujisaki, Motoyuki Shiga, Kei Moritsugu, and Akinori Kidera. Multiscale enhanced path sampling based on the Onsager–Machlup action: Application to a model polymer. *The Journal of Chemical Physics*, 139(5):054117, August 2013. ISSN 0021-9606. doi: 10.1063/1.4817209.
- [155] Thomas E. III Gartner and Arathi Jayaraman. Modeling and Simulations of Polymers: A Roadmap. *Macromolecules*, 52(3):755–786, February 2019. ISSN 0024-9297. doi: 10.1021/acs.macromol.8b01836.
- [156] Tobias Grafke, Tobias Schäfer, and Eric Vanden-Eijnden. Sharp Asymptotic Estimates for Expectations, Probabilities, and Mean First Passage Times in Stochastic Systems with Small Noise. *arXiv:2103.04837 [cond-mat, physics:physics]*, March 2021.

- [157] Bruce Donald, Patrick Xavier, John Canny, and John Reif. Kinodynamic motion planning. *Journal of the ACM*, 40(5):1048–1066, November 1993. ISSN 0004-5411. doi: 10.1145/174147.174150.
- [158] B. R. Donald and P. G. Xavier. Provably good approximation algorithms for optimal kinodynamic planning for Cartesian robots and open-chain manipulators. *Algorithmica*, 14(6):480–530, December 1995. ISSN 1432-0541. doi: 10.1007/BF01586637.
- [159] Federico Renda, Frederic Boyer, Jorge Dias, and Lakmal Seneviratne. Discrete Cosserat Approach for Multi-Section Soft Robots Dynamics. *IEEE Transactions on Robotics*, 34(6):1518–1533, December 2018. ISSN 1552-3098, 1941-0468. doi: 10.1109/TRO.2018.2868815.
- [160] Stanislao Grazioso, Giuseppe Di Gironimo, and Bruno Siciliano. A Geometrically Exact Model for Soft Continuum Robots: The Finite Element Deformation Space Formulation. *Soft Robotics*, 6(6):790–811, December 2019. ISSN 2169-5180. doi: 10.1089/soro.2018.0047.
- [161] Federico Renda, Vito Cacucciolo, Jorge Dias, and Lakmal Seneviratne. Discrete Cosserat approach for soft robot dynamics: A new piece-wise constant strain model with torsion and shears. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5495–5502, October 2016. doi: 10.1109/IROS.2016.7759808.
- [162] Brandon Caasenbrood, Alexander Pogromsky, and Henk Nijmeijer. Energy-Shaping Controllers for Soft Robot Manipulators Through Port-Hamiltonian Cosserat Models. *SN Computer Science*, 3(6):494, September 2022. ISSN 2661-8907. doi: 10.1007/s42979-022-01373-w.
- [163] F. Boyer, M. Porez, and W. Khalil. Macro-continuous computed torque algorithm for a three-dimensional eel-like robot. *IEEE Transactions on Robotics*, 22(4):763–775, August 2006. ISSN 1941-0468. doi: 10.1109/TRO.2006.875492.
- [164] Frédéric Boyer, Shaukat Ali, and Mathieu Porez. Macrocontinuous Dynamics for Hyperredundant Robots: Application to Kinematic Locomotion Bioinspired by Elongated Body Animals. *IEEE Transactions on Robotics*, 28(2):303–317, April 2012. ISSN 1552-3098, 1941-0468. doi: 10.1109/TRO.2011.2171616.
- [165] Federico Renda, Michele Giorelli, Marcello Calisti, Matteo Cianchetti, and Cecilia Laschi. Dynamic Model of a Multibending Soft Robot Arm Driven by Cables. *IEEE Transactions on Robotics*, 30(5):1109–1122, October 2014. ISSN 1941-0468. doi: 10.1109/TRO.2014.2325992.

- [166] Alexander Verl, Alin Albu-Schäffer, Oliver Brock, and Annika Raatz. *Soft Robotics: Transferring Theory to Application*. Springer, March 2015. ISBN 978-3-662-44506-8.
- [167] Noel Naughton, Jiarui Sun, Arman Tekinalp, Tejaswin Parthasarathy, Girish Chowdhary, and Mattia Gazzola. Elastica: A Compliant Mechanics Environment for Soft Robotic Control. *IEEE Robotics and Automation Letters*, 6(2):3389–3396, April 2021. ISSN 2377-3766, 2377-3774. doi: 10.1109/LRA.2021.3063698.
- [168] Padmini Rangamani, Ayelet Benjamini, Ashutosh Agrawal, Berend Smit, David J. Steigmann, and George Oster. Small scale membrane mechanics. *Biomechanics and Modeling in Mechanobiology*, 13(4):697–711, August 2014. ISSN 1617-7940. doi: 10.1007/s10237-013-0528-6.
- [169] K. L. Sack, S. Skatulla, and C. Sansour. Biological tissue mechanics with fibres modelled as one-dimensional Cosserat continua. Applications to cardiac tissue. *International Journal of Solids and Structures*, 81:84–94, March 2016. ISSN 0020-7683. doi: 10.1016/j.ijsolstr.2015.11.009.
- [170] Xiaotian Zhang, Fan Kiat Chan, Tejaswin Parthasarathy, and Mattia Gazzola. Modeling and simulation of complex dynamic musculoskeletal architectures. *Nature Communications*, 10(1):4825, December 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-12759-5.
- [171] Abhrajit Laskar and R. Adhikari. Brownian microhydrodynamics of active filaments. *Soft Matter*, 11(47):9073–9085, November 2015. ISSN 1744-6848. doi: 10.1039/C5SM02021B.
- [172] Abhrajit Laskar and R. Adhikari. Filament actuation by an active colloid at low Reynolds number. *New Journal of Physics*, 19(3):033021, March 2017. ISSN 1367-2630. doi: 10.1088/1367-2630/aa5f80.
- [173] Bartosz Kaczmarski, Derek E. Moulton, Ellen Kuhl, and Alain Goriely. Active filaments I: Curvature and torsion generation. *Journal of the Mechanics and Physics of Solids*, page 104918, May 2022. ISSN 0022-5096. doi: 10.1016/j.jmps.2022.104918.
- [174] Ankita Pandey, P. B. Sunil Kumar, and R. Adhikari. Flow-induced nonequilibrium self-assembly in suspensions of stiff, apolar, active filaments. *Soft Matter*, 12(44):9068–9076, November 2016. ISSN 1744-6848. doi: 10.1039/C6SM02104B.
- [175] M. Gazzola, L. H. Dudte, A. G. McCormick, and L. Mahadevan. Forward and inverse problems in the mechanics of soft filaments. *Royal Society Open Science*, 5(6):171628, June 2018. ISSN 2054-5703. doi: 10.1098/rsos.171628.

- [176] Raymond E. Goldstein, Philip Nelson, Thomas Powers, and Udo Seifert. Front Propagation in the Pearlning Instability of Tubular Vesicles. *Journal de Physique II*, 6(5):767–796, May 1996. ISSN 1155-4312, 1286-4870. doi: 10.1051/jp2:1996210.
- [177] Raymond E. Goldstein, Thomas R. Powers, and Chris H. Wiggins. Viscous Nonlinear Dynamics of Twist and Writhe. *Physical Review Letters*, 80(23):5232–5235, June 1998. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.80.5232.
- [178] Raymond E. Goldstein and Stephen A. Langer. Nonlinear Dynamics of Stiff Polymers. *Physical Review Letters*, 75(6):1094–1097, August 1995. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.75.1094.
- [179] Ioannis Stefanou, Jean Sulem, and Hadrien Rattez. Cosserat Approach to Localization in Geomaterials. In George Z. Voyatzis, editor, *Handbook of Nonlocal Continuum Mechanics for Materials and Structures*, pages 1–25. Springer International Publishing, Cham, 2017. ISBN 978-3-319-22977-5. doi: 10.1007/978-3-319-22977-5_10-1.
- [180] Samuel Forest, Fabrice Barbe, and Georges Cailletaud. Cosserat modelling of size effects in the mechanical behaviour of polycrystals and multi-phase materials. *International Journal of Solids and Structures*, 37(46):7105–7126, November 2000. ISSN 0020-7683. doi: 10.1016/S0020-7683(99)00330-3.
- [181] Holm Altenbach and Victor A. Eremeyev. Cosserat Media. In Holm Altenbach and Victor A. Eremeyev, editors, *Generalized Continua from the Theory to Engineering Applications*, CISM International Centre for Mechanical Sciences, pages 65–130. Springer, Vienna, 2013. ISBN 978-3-7091-1371-4. doi: 10.1007/978-3-7091-1371-4_2.
- [182] Babak Ebrahimian, Ali Noorzad, and Mustafa I. Alsaleh. A numerical study on interface shearing of granular Cosserat materials. *European Journal of Environmental and Civil Engineering*, 25(13):2337–2369, November 2021. ISSN 1964-8189. doi: 10.1080/19648189.2019.1627249.
- [183] L. Srinivasa Mohan, Prabhu R. Nott, and K. Kesava Rao. A frictional Cosserat model for the flow of granular materials through a vertical channel. *Acta Mechanica*, 138(1):75–96, March 1999. ISSN 1619-6937. doi: 10.1007/BF01179543.
- [184] Hidetoshi Kotera, Muneyo Sawada, and Susumu Shima. Cosserat continuum theory to simulate microscopic rotation of magnetic powder in applied magnetic field. *International Journal of Mechanical Sciences*, 42(1):129–145, January 2000. ISSN 0020-7403. doi: 10.1016/S0020-7403(98)00108-8.

- [185] I. Stefanou, J. Sulem, and I. Vardoulakis. Three-dimensional Cosserat homogenization of masonry structures: Elasticity. *Acta Geotechnica*, 3(1):71–83, March 2008. ISSN 1861-1133. doi: 10.1007/s11440-007-0051-y.
- [186] Patrick R. Onck. Cosserat modeling of cellular solids. *Comptes Rendus Mécanique*, 330(11):717–722, November 2002. ISSN 1631-0721. doi: 10.1016/S1631-0721(02)01529-2.
- [187] D. Ieşan. Deformation of porous Cosserat elastic bars. *International Journal of Solids and Structures*, 48(3):573–583, February 2011. ISSN 0020-7683. doi: 10.1016/j.ijsolstr.2010.10.022.
- [188] H. C. Park and R. S. Lakes. Cosserat micromechanics of human bone: Strain redistribution by a hydration sensitive constituent. *Journal of Biomechanics*, 19(5):385–397, 1986. ISSN 0021-9290. doi: 10.1016/0021-9290(86)90015-1.
- [189] Marcelo Epstein and Manuel de León. Continuous Distributions of Inhomogeneities in Liquid-Crystal-Like Bodies. *Proceedings: Mathematical, Physical and Engineering Sciences*, 457(2014):2507–2520, 2001. ISSN 1364-5021.
- [190] Alain Goriely, Derek E. Moulton, and L. Angela Mihai. A Rod Theory for Liquid Crystalline Elastomers. *Journal of Elasticity*, January 2022. ISSN 0374-3535, 1573-2681. doi: 10.1007/s10659-021-09875-z.
- [191] Elena A. Ivanova. On a New Theory of the Cosserat Continuum with Applications in Electrodynamics. In Holm Altenbach, Svetlana Bauer, Victor A. Eremeyev, Gennadi I. Mikhasev, and Nikita F. Morozov, editors, *Recent Approaches in the Theory of Plates and Plate-Like Structures*, Advanced Structured Materials, pages 75–87. Springer International Publishing, Cham, 2022. ISBN 978-3-030-87185-7. doi: 10.1007/978-3-030-87185-7_7.
- [192] Gundhar Paria. A Unified Theory of Mechanics of Ferro-Magnetic Material from Cosserat Point of View. *Bulletin mathématique de la Société des Sciences Mathématiques de la République Socialiste de Roumanie*, 22 (70)(3):303–314, 1978. ISSN 0373-2908.
- [193] E. A. Ivanova. Modeling of physical fields by means of the Cosserat continuum. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, n/a(n/a):e202100333. ISSN 1521-4001. doi: 10.1002/zamm.202100333.
- [194] J. L. Ericksen. Conservation Laws for Liquid Crystals. *Transactions of the Society of Rheology*, 5(1):23–34, March 1961. ISSN 0038-0032. doi: 10.1122/1.548883.

- [195] Stuart S. Antman. The Special Cosserat Theory of Rods. In Stuart S. Antman, editor, *Nonlinear Problems of Elasticity*, Applied Mathematical Sciences, pages 259–324. Springer, New York, NY, 1995. ISBN 978-1-4757-4147-6. doi: 10.1007/978-1-4757-4147-6_8.
- [196] M. B. Rubin. Cosserat Rods. In M. B. Rubin, editor, *Cosserat Theories: Shells, Rods and Points*, Solid Mechanics and Its Applications, pages 191–310. Springer Netherlands, Dordrecht, 2000. ISBN 978-94-015-9379-3. doi: 10.1007/978-94-015-9379-3_5.
- [197] D. F. Parker. On the derivation of nonlinear rod theories from three-dimensional elasticity. *Zeitschrift für Angewandte Mathematik und Physik (ZAMP)*, 35(6):833–847, November 1984. ISSN 0044-2275. doi: 10.1007/BF00945447.
- [198] J.C. Simo and L. Vu-Quoc. A Geometrically-exact rod model incorporating shear and torsion-warping deformation. *International Journal of Solids and Structures*, 27(3):371–393, 1991. ISSN 00207683. doi: 10.1016/0020-7683(91)90089-X.
- [199] J. C. Simo and L. Vu-Quoc. A three-dimensional finite-strain rod model. part II: Computational aspects. *Computer Methods in Applied Mechanics and Engineering*, 58(1):79–116, October 1986. ISSN 0045-7825. doi: 10.1016/0045-7825(86)90079-4.
- [200] Jeanne N. Clelland. *From Frenet to Cartan: The Method of Moving Frames*. American Mathematical Society, 2017. ISBN 978-1-4704-3747-3.
- [201] Félix Klein. *Development of Mathematics in the 19th Century: Appendices, "Kleinian Mathematics from an Advanced Standpoint"*. Math Sci Press, 1979. ISBN 978-0-915692-28-6.
- [202] Charles-Michel Marle. On Henri Poincaré’s Note “Sur une forme nouvelle des équations de la Mécanique”. *Journal of Geometry and Symmetry in Physics*, 29(none):1–38, January 2013. ISSN 1312-5192, 1314-5673. doi: 10.7546/jgsp-29-2013-1-38.
- [203] Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. Princeton University Press, October 2017. ISBN 978-0-691-17779-3.
- [204] Jakob Schwichtenberg. *Physics from Symmetry*. Springer International Publishing, December 2017. ISBN 978-3-319-66630-3.
- [205] Brian Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. Springer, May 2015. ISBN 978-3-319-13467-3.
- [206] Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of Differential Geometry, Volume 2*. Wiley, February 1996. ISBN 978-0-471-15732-8.

- [207] Harry Levy. Review: Elie Cartan, La Méthode de Repère Mobile, La Théorie des Groupes Continus, et Les Espaces Généralisés. *Bulletin of the American Mathematical Society*, 41(11):774–774, November 1935. ISSN 0002-9904, 1936-881X.
- [208] Peter J. Olver. A Survey of Moving Frames. In Hongbo Li, Peter J. Olver, and Gerald Sommer, editors, *Computer Algebra and Geometric Algebra with Applications*, Lecture Notes in Computer Science, pages 105–138, Berlin, Heidelberg, 2005. Springer. ISBN 978-3-540-32119-4. doi: 10.1007/11499251_11.
- [209] Robert B. Gardner. *The Method of Equivalence and Its Applications*. Society for Industrial and Applied Mathematics, 1989. ISBN 978-0-89871-240-7.
- [210] G. G. Giusteri, E. Miglio, N. Parolini, M. Penati, and R. Zambetti. Simulation of viscoelastic Cosserat rods based on the geometrically exact dynamics of special Euclidean strands, October 2021.
- [211] Giulio G Giusteri. Importance and Effectiveness of Representing the Shapes of Cosserat Rods and Framed Curves as Paths in the Special Euclidean Algebra. page 23.
- [212] Darryl D Holm, Jerrold E Marsden, and Tudor S Ratiu. The Euler–Poincaré Equations and Semidirect Products with Applications to Continuum Theories. *Advances in Mathematics*, 137(1):1–81, July 1998. ISSN 0001-8708. doi: 10.1006/aima.1998.1721.
- [213] H. Cendra, D. D. Holm, J. E. Marsden, and T. S. Ratiu. Lagrangian Reduction, the Euler–Poincaré Equations, and Semidirect Products, May 1999.
- [214] Henri Poincaré. Sur une forme nouvelle des équations de la mécanique. *CR Acad. Sci.*, 132:369–371, 1901.
- [215] Rafal Wisniewski and Piotr Kulczycki. Euler-Poincaré reduction of externally forced rigid body motion. page 14.
- [216] Darryl D. Holm and Rossen I. Ivanov. Matrix G-strands. *Nonlinearity*, 27(6):1445–1469, May 2014. ISSN 0951-7715. doi: 10.1088/0951-7715/27/6/1445.
- [217] Charles-Michel Marle Marle. On Henri Poincaré’s Note ”Sur une forme nouvelle des équations de la Mécanique”. Technical report, Journal of Geometry and Symmetry in Physics, 2013.
- [218] Hernán Cendra, Darryl D. Holm, Mark J. W. Hoyle, and Jerrold E. Marsden. The Maxwell–Vlasov equations in Euler–Poincaré form. *Journal of Mathematical Physics*, 39(6):3138–3157, June 1998. ISSN 0022-2488, 1089-7658. doi: 10.1063/1.532244.

- [219] Darryl D Holm. Geometric Mechanics, Part II: Rotating, Translating and Rolling. page 304.
- [220] *Nonlinear Problems of Elasticity*, volume 107 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2005. ISBN 978-0-387-20880-0. doi: 10.1007/0-387-27649-1.
- [221] Tixian Wang, Udit Halder, Heng-Sheng Chang, Mattia Gazzola, and Prashant G. Mehta. Optimal Control of a Soft CyberOctopus Arm. In *2021 American Control Conference (ACC)*, pages 4757–4764, New Orleans, LA, USA, May 2021. IEEE. ISBN 978-1-66544-197-1. doi: 10.23919/ACC50511.2021.9483284.
- [222] Mohit Garg and Ajeet Kumar. A slender body theory for the motion of special Cosserat filaments in Stokes flow. *Mathematics and Mechanics of Solids*, page 10812865221083323, May 2022. ISSN 1081-2865. doi: 10.1177/10812865221083323.
- [223] John Till, Vincent Aloi, and Caleb Rucker. Real-time dynamics of soft and continuum robots based on Cosserat rod models. *The International Journal of Robotics Research*, 38(6):723–746, May 2019. ISSN 0278-3649. doi: 10.1177/0278364919842269.
- [224] Yavuz Basar and Dieter Weichert. *Nonlinear Continuum Mechanics of Solids: Fundamental Mathematical and Physical Concepts*. Springer Science & Business Media, November 2013. ISBN 978-3-662-04299-1.
- [225] L. D. Landau, Evgenij M. Lifšic, E. M. Lifshitz, A. M. Kosevich, and L. P. Pitaevskii. *Theory of Elasticity: Volume 7*. Elsevier, January 1986. ISBN 978-0-7506-2633-0.
- [226] D. Q. Cao and Robin W. Tucker. Nonlinear dynamics of elastic rods using the Cosserat theory: Modelling and simulation. *International Journal of Solids and Structures*, 45(2):460–477, January 2008. ISSN 0020-7683. doi: 10.1016/j.ijsolstr.2007.08.016.
- [227] Thomas R. Powers. Dynamics of filaments and membranes in a viscous fluid. *Reviews of Modern Physics*, 82(2):1607–1631, May 2010. ISSN 0034-6861, 1539-0756. doi: 10.1103/RevModPhys.82.1607.
- [228] R. P. Nordgren. On Computation of the Motion of Elastic Rods. *Journal of Applied Mechanics*, 41(3):777–780, September 1974. ISSN 0021-8936. doi: 10.1115/1.3423387.
- [229] Christophe Eloy and Eric Lauga. Kinematics of the Most Efficient Cilium. *Physical Review Letters*, 109(3):038101, July 2012. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.109.038101.

- [230] Anna-Karin Tornberg and Michael J. Shelley. Simulating the dynamics and interactions of flexible fibers in Stokes flows. *Journal of Computational Physics*, 196(1):8–40, May 2004. ISSN 00219991. doi: 10.1016/j.jcp.2003.10.017.
- [231] Kunitsugu Soda. Dynamics of Stiff Chains. I. Equation of Motion. *Journal of the Physical Society of Japan*, 35(3):866–870, September 1973. ISSN 0031-9015. doi: 10.1143/JPSJ.35.866.
- [232] Hidenori Hasimoto. A soliton on a vortex filament. *Journal of Fluid Mechanics*, 51(3):477–485, February 1972. ISSN 1469-7645, 0022-1120. doi: 10.1017/S0022112072002307.
- [233] Élie Cartan. La theorie des groupes finis et continus et la geometrie differentielle traitees par la methode du repere mobile. *Bulletin of the American Mathematical Society*, 44:598–601, 1938.
- [234] Richard L. Bishop. There is More than One Way to Frame a Curve. *The American Mathematical Monthly*, 82(3):246–251, 1975. ISSN 0002-9890. doi: 10.2307/2319846.
- [235] E. L. Mansfield and A. Rojo-Echeburúa. On the use of the rotation minimizing frame for variational systems with Euclidean symmetry. *Studies in Applied Mathematics*, 143(3):244–271, 2019. ISSN 1467-9590. doi: 10.1111/sapm.12275.
- [236] Daniel Carroll, Emek Kose, and Ivan Sterling. Improving Frenet’s Frame Using Bishop’s Frame. *Journal of Mathematics Research*, 5(4):p97, November 2013. ISSN 1916-9795. doi: 10.5539/jmr.v5n4p97.
- [237] J. M. Selig. Characterisation of Frenet–Serret and Bishop motions with applications to needle steering. *Robotica*, 31(6):981–992, September 2013. ISSN 0263-5747, 1469-8668. doi: 10.1017/S026357471300026X.
- [238] F. Frenet. Sur les courbes à double courbure. *Journal de Mathématiques Pures et Appliquées*, pages 437–447, 1852. ISSN 0021-7874.
- [239] J.-A. Serret. Sur quelques formules relatives à la théorie des courbes à double courbure. *Journal de Mathématiques Pures et Appliquées*, pages 193–207, 1851. ISSN 0021-7874.
- [240] Élie Cartan. *Geometry of Riemannian Spaces: Lie Groups; History, Frontiers and Applications Series*. Math Sci Press, 1983. ISBN 978-0-915692-34-7.
- [241] Mark Fels and Peter J. Olver. Moving Coframes: I. A Practical Algorithm. *Acta Applicandae Mathematica*, 51(2):161–213, April 1998. ISSN 1572-9036. doi: 10.1023/A:1005878210297.

- [242] Mark Fels and Peter J. Olver. Moving Coframes: II. Regularization and Theoretical Foundations. *Acta Applicandae Mathematica*, 55(2):127–208, January 1999. ISSN 1572-9036. doi: 10.1023/A:1006195823000.
- [243] Peter J Olver. Modern Developments in the Theory and Applications of Moving Frames. page 37.
- [244] Sergei V. Ketov. *Quantum Non-linear Sigma-Models: From Quantum Field Theory to Supersymmetry, Conformal Field Theory, Black Holes and Strings*. Springer Science & Business Media, March 2013. ISBN 978-3-662-04192-5.
- [245] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, Madan Rao, and R. Aditi Simha. Hydrodynamics of soft active matter. *Reviews of Modern Physics*, 85(3):1143–1189, July 2013. ISSN 0034-6861, 1539-0756. doi: 10.1103/RevModPhys.85.1143.
- [246] James P. Sethna. *Order Parameters, Broken Symmetry, and Topology*, pages 253–286. Oxford University PressOxford, second edition, January 2021. ISBN 978-0-19-886524-7 978-0-19-189761-0. doi: 10.1093/oso/9780198865247.003.0009.
- [247] P. M. Chaikin and T. C. Lubensky. *Principles of Condensed Matter Physics*. Cambridge University Press, Cambridge, 1995. ISBN 978-0-521-79450-3. doi: 10.1017/CBO9780511813467.
- [248] Javier Bonet and Richard D. Wood. *Nonlinear Continuum Mechanics for Finite Element Analysis*. Cambridge University Press, Cambridge, second edition, 2008. doi: 10.1017/CBO9780511755446.
- [249] Raj Kumar Manna and P. B. Sunil Kumar. Emergent topological phenomena in active polymeric fluids. *Soft Matter*, 15(3):477–486, January 2019. ISSN 1744-6848. doi: 10.1039/C8SM01981A.
- [250] Chiao-Peng Hsu, Alfredo Sciortino, Yu Alice de la Trobe, and Andreas R. Bausch. Activity-induced polar patterns of filaments gliding on a sphere. *Nature Communications*, 13(1):2579, May 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-30128-7.
- [251] Wolfgang Rindler. *Relativity: Special, General, and Cosmological*. Oxford University Press, 2001. ISBN 978-0-19-850835-9.
- [252] J. L. ERICKSEN. Liquid crystals and Cosserat surfaces. *The Quarterly Journal of Mechanics and Applied Mathematics*, 27(2):213–219, May 1974. ISSN 0033-5614. doi: 10.1093/qjmam/27.2.213.

- [253] Gregory S. Chirikjian. The Stochastic Elastica and Excluded-Volume Perturbations of DNA Conformational Ensembles. *International journal of non-linear mechanics*, 43(10):1108–1120, December 2008. ISSN 0020-7462. doi: 10.1016/j.ijnonlinmec.2008.10.005.
- [254] M. G. Bawendi and Karl F. Freed. A Wiener integral model for stiff polymer chains. *The Journal of Chemical Physics*, 83(5):2491–2496, September 1985. ISSN 0021-9606. doi: 10.1063/1.449296.
- [255] Hagen Kleinert. *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*. World Scientific, 2009. ISBN 978-981-4273-57-2.
- [256] T. B. Liverpool and S. F. Edwards. Probability distribution of wormlike polymer loops. *The Journal of Chemical Physics*, 103(15):6716–6719, October 1995. ISSN 0021-9606. doi: 10.1063/1.470350.
- [257] Gregory S. Chirikjian. Stochastic Processes on Lie Groups. In Gregory S. Chirikjian, editor, *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*, Applied and Numerical Harmonic Analysis, pages 361–388. Birkhäuser, Boston, 2012. ISBN 978-0-8176-4944-9. doi: 10.1007/978-0-8176-4944-9_11.
- [258] Elena Celledoni, Håkon Marthinsen, and Brynjulf Owren. An introduction to Lie group integrators – basics, new developments and applications. *Journal of Computational Physics*, 257:1040–1061, January 2014. ISSN 0021-9991. doi: 10.1016/j.jcp.2012.12.031.
- [259] Brynjulf Owren. Lie group integrators. *arXiv:1601.04664 [math]*, January 2016.
- [260] Arieh Iserles, Hans Munthe-Kaas, Syvert Nørsett, and Antonella Zanna. Lie-group methods. *Acta Numerica*, 2005. doi: 10.1017/S0962492900002154.
- [261] Elena Celledoni, Ergys Çokaj, Andrea Leone, Davide Murari, and Brynjulf Owren. Lie Group integrators for mechanical systems. *International Journal of Computer Mathematics*, 99(1):58–88, January 2022. ISSN 0020-7160, 1029-0265. doi: 10.1080/00207160.2021.1966772.
- [262] Hans Munthe-Kaas. High order Runge-Kutta methods on manifolds. *Applied Numerical Mathematics*, 29(1):115–127, January 1999. ISSN 0168-9274. doi: 10.1016/S0168-9274(98)00030-0.
- [263] C. J. Budd and A. Iserles. Geometric Integration: Numerical Solution of Differential Equations on Manifolds. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 357(1754):945–956, 1999. ISSN 1364-503X.

- [264] Kenth Engø and Stig Faltinsen. Numerical Integration of Lie–Poisson Systems While Preserving Coadjoint Orbits and Energy. *SIAM Journal on Numerical Analysis*, 39(1):128–145, January 2001. ISSN 0036-1429. doi: 10.1137/S0036142999364212.
- [265] Jack K. Hale. *Ordinary Differential Equations*. Courier Corporation, January 2009. ISBN 978-0-486-47211-9.
- [266] Wulf Rossmann. *Lie Groups: An Introduction Through Linear Groups*. Oxford University Press, 2006. ISBN 978-0-19-920251-5.
- [267] Jose-Luis Blanco. A tutorial on SE(3) transformation parameterizations and on-manifold optimization. *Technical report*, page 58.
- [268] Hans Munthe-Kaas. Runge-Kutta methods on Lie groups. *BIT Numerical Mathematics*, 38(1):92–111, March 1998. ISSN 1572-9125. doi: 10.1007/BF02510919.
- [269] Uri M. Ascher and Linda R. Petzold. *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM, August 1998. ISBN 978-0-89871-412-8.
- [270] S. Levit and U. Smilansky. A Theorem on Infinite Products of Eigenvalues of Sturm-Liouville Type Operators. *Proceedings of the American Mathematical Society*, 65(2):299–302, 1977. ISSN 0002-9939. doi: 10.2307/2041911.
- [271] Gerald V. Dunne. Functional Determinants in Quantum Field Theory. *Journal of Physics A: Mathematical and Theoretical*, 41(30):304006, August 2008. ISSN 1751-8113, 1751-8121. doi: 10.1088/1751-8113/41/30/304006.
- [272] Wilhelm Magnus. On the exponential solution of differential equations for a linear operator. *Communications on Pure and Applied Mathematics*, 7(4):649–673, 1954. ISSN 1097-0312. doi: 10.1002/cpa.3160070404.
- [273] Aaron Meurer, Christopher P. Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B. Kirpichev, Matthew Rocklin, AMiT Kumar, Sergiu Ivanov, Jason K. Moore, and Sartaj Singh. SymPy: Symbolic computing in Python. *PeerJ Computer Science*, 3:e103, 2017.

Appendix A

MCMC benchmark models

Here we define the Langevin model systems considered in Ch. 2 and Ch. 3. We begin by defining a non-dimensionalised formulation of the Langevin equation with additive noise

$$d\mathbf{X} = \mu \mathbf{F} dt + \sqrt{2D} d\mathbf{W}. \quad (\text{A.1})$$

where μ is the mobility constant and D the diffusion constant. We write the latter as $D = \mu k_B \theta$, where k_B is the Boltzmann constant θ is the temperature. We assume that the force is the gradient of a potential energy function $U : \mathbb{R}^d \rightarrow \mathbb{R}$ as $\mathbf{F} = -\nabla \cdot U$.

Let $T_0 = \frac{L^2}{k_B \theta_0 \mu}$ be a time-scale, θ_0 a temperature-scale, and L_0 a characteristic length-scale of the potential. We set $t = T_0 \tilde{t}$, $\theta = \theta_0 \tilde{\theta}$, $\mathbf{X} = L_0 \tilde{\mathbf{X}}$, $\mathbf{F} = \frac{\tilde{U}_0}{L_0} \tilde{\mathbf{F}}$ and $\mathbf{W} = \sqrt{T_0} \tilde{\mathbf{W}}$. T_0 is the typical diffusion time-scale at temperature θ_0 . We get

$$d\tilde{\mathbf{X}} = \tilde{U}_0 \tilde{\mathbf{F}} d\tilde{t} + \sqrt{2\tilde{\theta}} d\tilde{\mathbf{W}}. \quad (\text{A.2})$$

where all quantities in the equation are now non-dimensional, and where $\tilde{U}_0 = \frac{U_0}{k_B \theta_0}$ is the ratio of the energetic well-depth U_0 and the thermal energy at temperature θ_0 . We set $\tilde{U}_0 = 1$, such that $\theta = \theta_0$ corresponds to the temperature at which $U_0 = k_B \theta_0$. In other words, the characteristic energy scale of the potential is equal to the thermal energy at temperature θ_0 . Finally, the non-diemsionalised Langevin equation is

$$d\tilde{\mathbf{X}} = \tilde{\mathbf{F}} d\tilde{t} + \sqrt{2\tilde{\theta}} d\tilde{\mathbf{W}}. \quad (\text{A.3})$$

A.1 The asymmetric double-well system

Consider a 1-dimensional Langevin system with a quartic potential

$$U(x) = U_0 \left(1 + a \frac{x}{L_0} + b \left(\frac{x}{L_0} \right)^2 + x \left(\frac{x}{L_0} \right)^3 + d \left(\frac{x}{L_0} \right)^4 \right). \quad (\text{A.4})$$

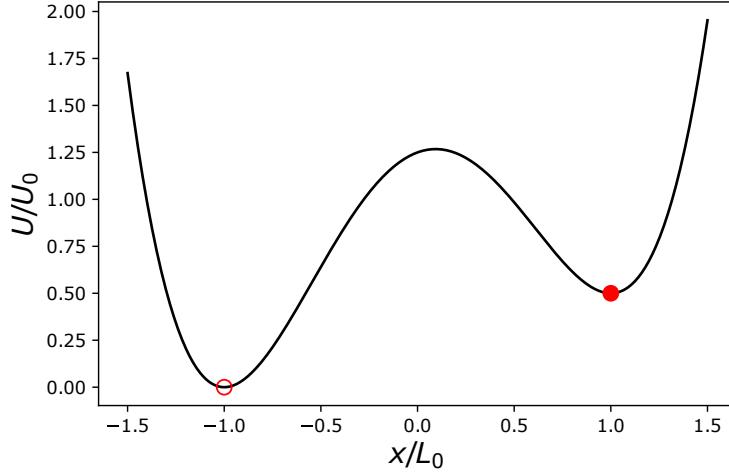


Figure A.1: The asymmetric double-well potential, with minima at $x/L_0 = -1$ (hollow red circle) and $x/L_0 = 1$ (filled red circle). See App. A.1 for its definition.

We impose the following conditions

1. $U'(-L_0) = U'(L_0) = 0$. Extrema at $x = -L_0$ and $x = L_0$.
2. $U(-L_0) = 0$ and $U(L_0) = \Delta U$. The energy cost of moving from the left-most extremum to the right-most extremum is ΔU .

Imposing the conditions on Eq.A.4, we get

$$U(x) = \left(\left(\frac{x}{L_0} - 1 \right)^2 - \frac{1}{4} \frac{\Delta U}{U_0} \left(\frac{x}{L_0} - 2 \right) \right) \left(\frac{x}{L_0} + 1 \right)^2. \quad (\text{A.5})$$

Equation A.5 defines an *asymmetric* variant of a standard quartic double-well system, with minimii at $\frac{x}{L_0} = -1$ and $\frac{x}{L_0} = 1$, and a maximum at $\frac{x}{L_0} = \frac{3\Delta U}{16U_0}$. See Fig. A.1 for a depiction of the potential energy landscape. In non-dimensionalised form, the potential is

$$\tilde{U}(\tilde{x}) = (\tilde{x} - 1)^2 - \frac{1}{4} \frac{\Delta U}{U_0} (\tilde{x} - 2)(\tilde{x} + 1)^2. \quad (\text{A.6})$$

where $\tilde{x} = x/L$. In all numerical simulations, we have set $\frac{\Delta U}{U_0} = 1/2$.

A.2 The switch system

Here we define the *switch* system, and its non-conservative variant used in Ch. 3. We consider a 2-dimensional Langevin system with a Sombrero-type potential $U : \mathbb{R}^2 \rightarrow \mathbb{R}$. Let $\Gamma = \{(x, y) \mid \sqrt{x^2 + y^2} = 1\}$ be the circle centred around the origin, and let Γ^+ and Γ^- signify the upper and lower semi-circle respectively. We will construct the potential such that Γ coincides with a manifold of potential minima, and such that the perpendicular

curvature of the potential along Γ^+ is larger than along Γ^- . We start with a radial quartic potential of the form

$$U(x_1, x_2) = U_r(r(x_1, x_2)) \quad (\text{A.7})$$

$$U_r(r) = U_0 \left(\frac{r}{L_0} - 1 \right) \left(1 + a \frac{r}{L_0} + b \left(\frac{r}{L_0} \right)^2 + c \left(\frac{r}{L_0} \right)^3 \right)$$

where L_0 is the length-scale of the system, U_0 will be the value of the potential at the local maximum $r = 0$, and $a, b, c \in \mathbb{R}$ will be specified below. We will henceforth suppress the argument of the radial coordinate function $r(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$. We impose the following conditions on the potential to fix a, b and c :

1. $U'_r(0) = 0$. The origin is an extremum.
2. $U'_r(L_0) = 0$. Γ is an extremum of the potential.
3. $U''_r(L_0) = k$. The curvature along Γ is k .

We get

$$U_r(r) = \frac{1}{2} \left(\frac{r}{L_0} - 1 \right)^2 \left[L_0^2 k \left(\frac{r}{L_0} \right)^2 - 2U_0 \left(\frac{r}{L_0} - 1 \right) \left(3 \frac{r}{L_0} + 1 \right) \right] \quad (\text{A.8})$$

which is, in the non-dimensionalised form

$$\tilde{U}_r(\tilde{r}) = \frac{1}{2} (\tilde{r} - 1)^2 \left[L_0^2 \tilde{k} \tilde{r}^2 - 2(\tilde{r} - 1)(3\tilde{r} + 1) \right]. \quad (\text{A.9})$$

where $\tilde{r} = r/L$ and $\tilde{k} = \frac{k}{U_0}$. In order to ensure that the potential has a Sombrero-like form, we must further have that the potential is confining, which is equivalent to $\lim_{r \rightarrow \infty} U_r(r) = \infty$, which implies that $6U_0 \leq L_0^2 k$. We now introduce an angular dependence in the curvature. We set

$$L_0^2 k(\phi) = 6U_0(1 + 2h(\phi)) \quad (\text{A.10})$$

where $\phi = \phi(x_1, x_2)$ is the angle of (x_1, x_2) in polar coordinates so that $x_1 = \cos(\phi)$, $x_2 = \sin(\phi)$, and where

$$h(\phi) = \frac{1}{4} (\xi_2 + \xi_1 + (\xi_2 - \xi_1) \sin \phi) \quad (\text{A.11})$$

where $\xi_2 > \xi_1$, and where $h(\phi) \in [\xi_1, \xi_2]$ satisfies $h(-\pi/2) = \xi_1$ and $h(\pi/2) = \xi_2$. Eq. (A.10) is constructed so that the perpendicular curvature of Γ^+ is larger than that of Γ^- . The drift of the system is now given by $\mathbf{F} = -\nabla U$. This concludes the definition of the switch system.

We also consider a non-conservative variant of the switch system. We introduce an additional non-conservative force $\mathbf{F}^a = -\eta \hat{\phi}$ for which the work done in a displacement

$d\mathbf{x} = dr\hat{\mathbf{r}} + rd\phi\hat{\boldsymbol{\phi}}$ is $dW = \mathbf{F}^a \cdot d\mathbf{x} = \eta r d\phi$. This force energetically biases the upper transition channel Γ^+ . The total force is thus $\mathbf{F} = -\nabla U + \mathbf{F}^a$.

In the numerical experiments presented in the main text, we used the Itô Langevin equation

$$d\mathbf{X} = \mu \mathbf{F} dt + \sqrt{2\mu k_B \theta} d\mathbf{W}. \quad (\text{A.12})$$

For the numerical experiments in the main text we use $\tilde{U}_0 = 1$, which means that $\tilde{\theta} = 1$ corresponds to a temperature such that $k_B\theta = U_0$. We also set $\xi_1 = 0$ and $\xi_2 = 2$. To compare the gradient force $-\nabla \cdot U$ with \mathbf{F}^a we also introduce $f_{\text{eq}} = U_0/L$, which is the characteristic force strength of the gradient force.

Appendix B

Calculating the regularised normalisation constants of infinite-dimensional Gaussian measures

The regularised normalisation constants of Gaussian measures defined on functional spaces can be found by computing the determinants of their covariance operators. Equivalently, they can be found by computing the determinant of their precision operator, which is the inverse of the covariance operator. As for finite-dimensional linear operators, determinants of differential operators can be found by computing the product of their eigenvalues, but as the numerical computation of the spectrum of a linear operator is expensive, this is in general not feasible in this setting. For a specific class of linear operators, their regularised normalisation constants can be computed using the Gelfand-Yaglom theorem (GYT) [33, 270, 271], by solving a system of ODEs. Here we will first recount the GYT, and then derive an extension of theorem that can be used for the precision operators of the semi-classical expansions of the laws of general Itô processes with additive noise.

Let the linear operators

$$\mathcal{L} = \frac{d}{dt} \left(P \frac{d}{dt} \right) - R \quad (\text{B.1})$$

and

$$\mathcal{L}_0 = \frac{d}{dt} \left(P \frac{d}{dt} \right) \quad (\text{B.2})$$

be defined for $0 \leq t \leq T$, and where $P(t) \in \mathbb{R}^{d \times d}$ is a symmetric positive-definite matrix function and $R(t) \in \mathbb{R}^{d \times d}$ is a symmetric matrix function. Let $\gamma^{(k)}$ and $\mathbf{u}^{(k)}(t; \alpha)$ be the eigenvalues and eigenfunctions of \mathcal{L} , which are solutions to the boundary value problem

$$\mathcal{L}\mathbf{u}^{(k)}(t) = \gamma^{(k)}\mathbf{u}^{(k)}(t) \quad (\text{B.3})$$

where $\mathbf{u}^{(k)}(0) = \mathbf{u}^{(k)}(T) = 0$. Similarly, let $\mathbf{u}_0^{(k)}(t)$ and $\gamma_0^{(k)}$ be the eigenfunctions and eigenvalues of \mathcal{L}_0 . Then the *functional determinant* of \mathcal{L} is defined in regularised form as

$$\frac{\det \mathcal{L}}{\det \mathcal{L}_0} = \prod_{k=1}^{\infty} \frac{\gamma^{(k)}}{\gamma_0^{(k)}}. \quad (\text{B.4})$$

As the spectrum of Eq. B.1 is unknown, and numerically expensive to compute, a much more efficient way of computing Eq. B.4 is via the GYT. The GYT states that the functional determinant can be expressed as

$$\left| \frac{\det \mathcal{L}}{\det \mathcal{L}_0} \right| = \left| \frac{\det [Y(T)]}{\det [Y_0(T)]} \right| \quad (\text{B.5})$$

where $Y(t) \in \mathbb{R}^{d \times x}$ with components $Y_{ij}(t) = y_i^{(j)}(t)$, where the $\mathbf{y}^{(j)}(t)$ are solutions to the d second-order ODEs with initial conditions

$$\mathcal{L}\mathbf{y}^{(j)}(t) = 0 \quad (\text{B.6})$$

$$\mathbf{y}^{(j)}(0) = 0 \quad (\text{B.7})$$

$$\frac{d}{dt} y_i^{(j)}(0) = \delta_{ij}. \quad (\text{B.8})$$

and where the matrix $Y_0(t) \in \mathbb{R}^{d \times d}$ is defined similarly, but with \mathcal{L} in Eq. B.6 replaced with \mathcal{L}_0 . We now present a generalisation of the GYT that allows for linear operators of the form

$$\mathcal{L} = \frac{d^2}{dt^2} + U \frac{d}{dt} + R \quad (\text{B.9})$$

where $0 \leq t \leq T$, and where $U(t), R(t) \in \mathbb{R}^{d \times d}$ are anti-symmetric and symmetric matrix functions respectively. This extended GYT will make the theorem applicable to compute the regularised normalisation constants of the precision operators discussed in Sec. 2.4.1.

We define the linear operator \mathcal{G} which acts on vector functions as

$$\mathcal{G}\mathbf{y} = G\mathbf{y} \quad (\text{B.10})$$

where $G(t)$ is a matrix function that solves the equation

$$\dot{G} = -\frac{1}{2}UG. \quad (\text{B.11})$$

As $U(t)$ is anti-symmetric we have $U(t) \in \mathfrak{so}(d)$, where the latter is the Lie algebra of the special orthogonal group $SO(d)$. Then Eq. B.11 is the exponential mapping from the Lie

algebra to the Lie group, therefore $G \in SO(d)$. We define the linear operator

$$\tilde{\mathcal{L}} = \mathcal{G}^{-1} \mathcal{L} \mathcal{G} = \frac{d}{dt^2} + G^T \left(R - \frac{1}{2} \dot{U} - \frac{1}{4} U^2 \right) G \quad (\text{B.12})$$

where we have used $G^{-1} = G^T$. Equation B.12 is of the form Eq. B.1, as the second term is a symmetric matrix, and we can therefore compute $\det \tilde{\mathcal{L}}$ using the GYT. As for any two operators \mathcal{A} and \mathcal{B} , we have that $\det \mathcal{A}\mathcal{B} = \det \mathcal{A} \det \mathcal{B}$, and $\det \mathcal{A}^{-1} = 1/\det \mathcal{A}$, and we therefore have that $\det \mathcal{L} = \det \tilde{\mathcal{L}}$. The functional determinant $\det \mathcal{L}$ can thus be computed by first solving Eq. B.11, and then constructing $\tilde{\mathcal{L}}$ using Eq. B.12, and finally using the GYT to compute $\det \tilde{\mathcal{L}}$.

The above can be applied to compute the regularised normalisation constants of the precision operators derived in Sec. 2.4.1, which were the result the semi-classical approximations of the path measures of general Itô diffusions with additive noise. The precision matrix is of the form

$$\mathcal{H}^{[\bar{x}]} = -\frac{1}{D} \frac{d}{dt^2} + 2A \frac{d}{dt} + B \quad (\text{B.13})$$

We can compute the regularised normalisation constant of $\mathcal{H}^{[\bar{x}]}$ by letting $P(t) = -\frac{1}{D} \mathbb{I}_d$, $U = -2DA$ and $R = B$ and applying the extended GYT. In the case of gradient dynamics, when $A = 0$, the original GYT can be readily applied by using $P(t) = -\frac{1}{D} \mathbb{I}_d$ and $R(t) = -B(t)$, where \mathbb{I}_d is the identity matrix.

Appendix C

Reconstructing the spatio-temporal configuration

To reconstruct the spatio-temporal configuration $\Phi(t, \mathbf{u})$ of a system from its spatial reconstruction fields $X_\alpha(t, \mathbf{u})$ and generalised velocity field $N(t, \mathbf{u})$, where $\mathbf{u} \in M$, we must solve the equation

$$\Phi^{-1} d\Phi = \xi \quad (\text{C.1})$$

where $\xi = Ndt + X_\alpha du^\alpha$. Given some initial condition $\Phi(\bar{t}, \mathbf{u}_0)$, we can find $\Phi(\bar{t}, \mathbf{u})$ by integrating Eq. C.1 for fixed \bar{t} along some curve $\gamma : [0, 1] \rightarrow M$ that satisfies $\gamma(0) = \mathbf{u}_0$ and $\gamma(1) = \mathbf{u}$. If X_α satisfy spatial integrability, the result of this integration is independent of the shape of the path γ .

A numerical scheme for reconstructing the spatio-temporal configuration can thus be devised by repeatedly integrating Eq. C.1 along curves that trace out Φ along the desired points $\mathbf{u} \in M$. In the following section we will show such a scheme explicitly for the case of the Cosserat rod, and the procedure can be readily generalised to higher-dimensional systems.

C.1 Numerical algorithm for reconstructing the Cosserat rod

Here we describe how to reconstruction the spatio-temporal configuration of a Cosserat rod $\Phi(t, u)$ from the spatial reconstruction field $X(t, u)$ and generalised velocity field $N(t, u)$. These are related as Eq. C.1, where $\xi = Xdu + Ndt$, from which we have

$$\Phi' = \Phi X \quad (\text{C.2a})$$

$$\dot{\Phi} = \Phi N. \quad (\text{C.2b})$$

Formally, these have solutions

$$\Phi_{ij}(u, \bar{t}) = \Phi_{ik}(0, \bar{t}) \mathcal{U} \left\{ e^{\int_0^u X(u, \bar{t}) du} \right\}_{kj} \quad (\text{C.3a})$$

$$\Phi_{ij}(\bar{u}, t) = \Phi_{ik}(\bar{u}, 0) \mathcal{T} \left\{ e^{\int_0^t N(\bar{u}, t) dt} \right\}_{kj} \quad (\text{C.3b})$$

where $\mathcal{U}(\cdot)$ and $\mathcal{T}(\cdot)$ signifies the spatial and temporal time-ordered integral respectively. Numerically, these can be efficiently approximated by discretisation, and incrementally solved using the *Magnus expansion* [272]. Eq. C.2a and Eq. C.2b can therefore be used together to trace out \mathcal{F} from a single initial condition $\Phi(u_0, t_0)$. Due to the integrability of ξ , the value $\Phi(t, u)$ at (t, u) is invariant with respect to which particular path in (t, u) -space is taken. In light of this there are theoretically an infinite number of reconstruction schemes. Below we propose what we find to be the most convenient reconstruction scheme.

We discretise time and space as $u_\alpha = \alpha \Delta u$, $\alpha = 0, \dots, N_u$, such that $u_0 = t_0 = 0$ and $t_\beta = \beta \Delta t$, $\beta = 0, \dots, N_t$ and $u_{N_u} = L_0$ and $t_{N_t} = T$. We define $\Phi^{\alpha, \beta} \approx \Phi(\alpha \Delta u, \beta \Delta t)$, $X^{\alpha, \beta} \approx X(\alpha \Delta u, \beta \Delta t)$ and $Y^{\alpha, \beta} \approx Y(\alpha \Delta u, \beta \Delta t)$.

$$\Phi_{ij}^{0, \beta+1} = \Phi_{ik}^{0, \beta} \exp_{SE(3)} (\Delta t N^{0, \beta})_{kj}, \quad \Phi^{0, 0} = \Phi^{(i)} \quad (\text{C.4a})$$

$$\Phi_{ij}^{\alpha+1, \beta} = \Phi_{ik}^{\alpha, \beta} \exp_{SE(3)} (\Delta u X^{\alpha, \beta})_{kj} \quad (\text{C.4b})$$

with the initial condition $\mathcal{F}^{0, 0} = \mathcal{F}^{(i)}$. The matrix exponential has a closed-form analytical formula given by

$$\exp_{SE(3)}(H) = \begin{pmatrix} 1 & \vec{0}^T \\ \mathcal{B}(\hat{m})\vec{v} & \exp_{SO(3)}(\hat{m}) \end{pmatrix} \quad (\text{C.5})$$

for $H = \{\vec{v}; \vec{m}\} \in \mathfrak{se}(3)$ and

$$\mathcal{B}(\hat{m}) = \mathbb{1}_3 + \frac{1 - \cos |\vec{m}|}{|\vec{m}|^2} \hat{m} + \frac{|\vec{m}| - \sin |\vec{m}|}{|\vec{m}|^3} \hat{m}^2 \quad (\text{C.6a})$$

$$\exp_{SO(3)}(\hat{m}) = \mathbb{1}_3 + \frac{\sin |\vec{m}|}{|\vec{m}|} \hat{m} + \frac{1 - \cos |\vec{m}|}{|\vec{m}|^2} \hat{m}^2 \quad (\text{C.6b})$$

where $\exp_{SO(3)}(\hat{m}) \in SO(3)$ is an element of the special orthogonal group in 3 dimensions. As Eq. C.6a and Eq. C.6b are ill-conditioned for small $|\vec{m}|$, in numerics it is often prudent to Taylor expand to at least $O(|\vec{m}|^6)$ when $|\vec{m}| < 0.1$ [210].

A benefit of Eq. C.4 as a numerical solution schema is that it allows for flexibility in terms of when Eq. C.4b is evaluated. For example, in simulations we use a temporal discretisation Δt , but we can choose to only reconstruct the Cosserat rod at intervals $n \Delta t$, for some integer n .

Appendix D

Detailed derivation of the $SE(3)$ short-time propagator

Here we provide detailed derivations of the results in Sec. 7.3. Let $Y = \{\vec{a}; \vec{m}\} \in \mathfrak{se}(3)$, $N = \{\vec{V}; \vec{\Omega}\}$ and $\bar{\Delta}t = \Delta t - \Delta t'$, then

$$\begin{aligned} \mathcal{E}_{SE(3)}(\Delta t, N, Y) &= \int_0^{\Delta t} \exp_{SE(3)}(-\bar{\Delta}t N) Y \exp_{SE(3)}(\bar{\Delta}t N) d\Delta t' \\ &= \begin{pmatrix} 0 & \vec{0}^T \\ \textcircled{1} + \vec{r}(\Delta t, \vec{m}, \vec{V}) & \textcircled{2} \end{pmatrix} \end{aligned} \quad (\text{D.1})$$

where we used Eq. 7.26, and where

$$\textcircled{1} = \int_0^{\Delta t} \exp_{SO(3)}(-\bar{\Delta}t \hat{\Omega}) \vec{a} d\Delta t' \quad (\text{D.2a})$$

$$\textcircled{2} = \int_0^{\Delta t} \exp_{SO(3)}(-\bar{\Delta}t \hat{\Omega}) \hat{m} \exp_{SO(3)}(\bar{\Delta}t \hat{\Omega}) d\Delta t' \quad (\text{D.2b})$$

$$\vec{r}(\Delta t, N, Y) = \int_0^{\Delta t} \bar{\Delta}t \exp_{SO(3)}(-\bar{\Delta}t \hat{\Omega}) \hat{m} \mathcal{B}(\bar{\Delta}t \hat{\Omega}) \vec{V} d\Delta t'. \quad (\text{D.2c})$$

From the results of Sec. 7.2 we can make the identifications

$$\textcircled{1} = \Delta t \mathcal{B}(-\Delta t \hat{\Omega}) \vec{a} \quad (\text{D.3a})$$

$$\textcircled{2} = \mathcal{E}_{SO(3)}(\Delta t, \hat{\Omega}, \hat{m}). \quad (\text{D.3b})$$

where $\mathcal{B}(\hat{m})$ was defined in Eq. 7.24. It remains to evaluate $\vec{r}(\Delta t, N, Y)$. Let $\alpha = \Delta t'/\Delta t$, $\psi = \Delta t |\vec{\Omega}|$ and $\bar{\Omega} = \hat{\Omega}/|\vec{\Omega}|$ such that $(\Delta t - \Delta t') \hat{\Omega} = (1 - \alpha) \psi \bar{\Omega}$. We rewrite $\vec{r}(\Delta t, N, Y)$ as

$$\vec{r}(\Delta t, N, Y) = \int_0^1 \Delta t^2 (1 - \alpha) \exp_{SO(3)}(-(1 - \alpha) \psi \bar{\Omega}) \hat{m} \mathcal{B}((1 - \alpha) \psi \bar{\Omega}) \vec{V} d\alpha. \quad (\text{D.4})$$

We now evaluate the integrand explicitly, we get

$$\Delta t^2(1-\alpha)\exp_{SO(3)}(-(1-\alpha)\psi\bar{\Omega})\hat{m}\mathcal{B}((1-\alpha)\psi\bar{\Omega}) = \Delta t^2 \sum_{i,j=1}^3 P_{ij}\bar{\Omega}^{i-1}\hat{m}\bar{\Omega}^{j-1} \quad (\text{D.5})$$

where

$$P = \begin{pmatrix} \bar{\alpha} & \frac{\cos(\bar{\alpha}\psi)-1}{\psi} & \bar{\alpha} - \frac{\sin(\bar{\alpha}\psi)}{\psi} \\ \bar{\alpha}\sin(\bar{\alpha}\psi) & \frac{(\cos(\bar{\alpha}\psi)-1)\sin(\bar{\alpha}\psi)}{\psi} & \frac{(\bar{\alpha}\psi-\sin(\bar{\alpha}\psi))\sin(\bar{\alpha}\psi)}{\psi} \\ \bar{\alpha}(1-\cos(\bar{\alpha}\psi)) & -\frac{(\cos(\bar{\alpha}\psi)-1)^2}{\psi} & \bar{\alpha}(1-\cos(\bar{\alpha}\psi)) - \frac{2\sin(\bar{\alpha}\psi)+\sin(2\bar{\alpha}\psi)}{2\psi} \end{pmatrix} \quad (\text{D.6})$$

where the i th and j th indices of P_{ij} signify rows and columns respectively and $\bar{\alpha} = 1 - \alpha$. We can now integrate each component p_{ij} individually. We define

$$R = \psi^2 \int_0^1 d\alpha P = \begin{pmatrix} \frac{\psi^2}{2} & -\psi + \sin(\psi) & \frac{\psi^2}{2} + \cos(\psi) - 1 \\ -\psi \cos(\psi) + \sin(\psi) & -\frac{(\cos(\psi)-1)^2}{2} & -\psi \cos(\psi) - \frac{\psi}{2} + \sin(\psi) + \frac{\sin(2\psi)}{4} \\ \frac{\psi^2}{2} - \psi \sin(\psi) - \cos(\psi) + 1 & -\frac{3\psi}{2} + 2 \sin(\psi) - \frac{\sin(2\psi)}{4} & \frac{(\psi-\sin(\psi))^2}{2} \end{pmatrix}, \quad (\text{D.7})$$

such that

$$\vec{r}(\Delta t, N, \vec{m}) = \frac{1}{|\vec{\Omega}|^2} \sum_{i,j=1}^3 R_{ij}\bar{\Omega}^{i-1}\hat{m}\bar{\Omega}^{j-1}\vec{V}. \quad (\text{D.8})$$

We made use of the symbolical manipulation package *SymPy* [273] in order to find the expressions for P and R . We will now express the short-term propagator Eq. 7.18 in terms of the sub-matrices of $X_\alpha = \{\vec{\theta}_\alpha, \vec{\pi}_\alpha\}$, $N = \{\vec{V}; \vec{\Omega}\}$ $S = \{\vec{P}; \vec{L}\}^*$ and $Q = \{\vec{F}; \vec{M}\}^*$. Firstly we define

$$X_\alpha(t_0, \mathbf{u}) := X_{0,\alpha} = \{\vec{\theta}_{\alpha,0}; \vec{\pi}_{\alpha,0}\} \quad (\text{D.9a})$$

$$N(t_0, \mathbf{u}) := N_0 = \{\vec{V}_0; \vec{\Omega}_0\} \quad (\text{D.9b})$$

$$S(t_0, \mathbf{u}) := S_0 = \{\vec{P}_0; \vec{L}_0\}^* \quad (\text{D.9c})$$

$$F_\alpha(t_0, \mathbf{u}) := F_{0,\alpha} = \{\vec{f}_{\alpha,0}^1; \vec{f}_{\alpha,0}^2\} \quad (\text{D.9d})$$

$$H(t_0, \mathbf{u}) := H_0 = \{\vec{h}_0^1; \vec{h}_0^2\}^* \quad (\text{D.9e})$$

where F_α and H were defined in Eq. 7.10 using $Q^\alpha = \{\vec{F}^\alpha; \vec{M}^\alpha\}^*$, and

$$U_0^1 = \exp_{SO(3)}(-\Delta t \hat{\Omega}_0) \quad (\text{D.10a})$$

$$U_0^2 = \mathcal{B}(-\Delta t \hat{\Omega}_0) \quad (\text{D.10b})$$

$$\vec{s}_0(\vec{m}) = \Delta t(U_0^1 \vec{m}) \times (U_0^2 \vec{V}_0) \quad (\text{D.10c})$$

where \times denotes the cross product. Then we have that

$$\exp_{SE(3)}(-\Delta t N) X_{0,\alpha} \exp_{SE(3)}(\Delta t N) = \begin{pmatrix} 0 & \vec{0}^T \\ U_0^1 \vec{\theta}_0 + \vec{s}_0(\vec{\pi}_{0,\alpha}) & \widehat{U_0^1 \vec{\pi}_{0,\alpha}} \end{pmatrix} \quad (\text{D.11a})$$

$$\exp_{SE(3)}(-\Delta t N) S_0^T \exp_{SE(3)}(\Delta t N) = \begin{pmatrix} 0 & \vec{0}^T \\ U_0^1 \vec{P}_0 + \vec{s}_0(\vec{P}_0) & \widehat{U_0^1 \vec{L}_0} \end{pmatrix}. \quad (\text{D.11b})$$

Then, using Eq. D.11 with Eq. D.1 we can expand Eq. 7.18 as

$$\vec{\theta}_\alpha(t_0 + \Delta t, \mathbf{u}) \approx U_0^1 \vec{\theta}_{0,\alpha} + \Delta t U_0^2 \vec{f}_{0,\alpha}^1 + \vec{s}_0(\vec{\pi}_{0,\alpha}) + \vec{r}(\Delta t, N_0, \vec{f}_{0,\alpha}^2) \quad (\text{D.12a})$$

$$\vec{\pi}_\alpha(t_0 + \Delta t, \mathbf{u}) \approx U_0^1 \vec{\pi}_{0,\alpha} + \mathcal{E}_{SO(3)}(\Delta t, \hat{\Omega}, \vec{f}_{0,\alpha}^2) \quad (\text{D.12b})$$

$$\vec{P}(t_0 + \Delta t, \mathbf{u}) \approx U_0^1 \vec{P}_0 + \Delta t U_0^2 \vec{h}_0^1 + \vec{s}_0(\vec{L}_0) + \vec{r}(\Delta t, N_0, \vec{h}_0^2) \quad (\text{D.12c})$$

$$\vec{L}(t_0 + \Delta t, \mathbf{u}) \approx U_0^1 \vec{L}_0 + \mathcal{E}_{SO(3)}(\Delta t, \hat{\Omega}, \vec{h}_0^2). \quad (\text{D.12d})$$

D.1 Simplified integrator for $SE(3)$ -configured systems

The kinodynamic equations of an $SE(3)$ -configured system, if expressed in terms of the sub-matrices of X_α and S , are

$$D_t \vec{\theta}_\alpha = D_\alpha \vec{V} \quad (\text{D.13a})$$

$$\partial_t \vec{\pi}_\alpha = D_\alpha \vec{\Omega} \quad (\text{D.13b})$$

$$D_t \vec{P} = D_\alpha \vec{F}^\alpha + \vec{f} \quad (\text{D.13c})$$

$$D_t \vec{L} = D_\alpha \vec{M}^\alpha + \vec{\theta}_\alpha \times \vec{F}^\alpha + \vec{m} \quad (\text{D.13d})$$

where $D_t = \partial_t + \hat{\Omega}$ and $D_\alpha = \partial_\alpha + \hat{\pi}$ and $\alpha = 1, \dots, d$. We can construct a simplified short-term propagator by noting that each individual equation in Eq. D.13 can be integrated using the propagator derived in Sec. 7.2. To see that we rewrite Eq. D.13 as

$$\partial_t \vec{\theta}_\alpha = G_\alpha + A \vec{\theta}_\alpha \quad (\text{D.14a})$$

$$\partial_t \vec{\pi}_\alpha = H_\alpha + A \vec{\pi}_\alpha \quad (\text{D.14b})$$

$$\partial_t \vec{P} = K + A \vec{P} \quad (\text{D.14c})$$

$$\partial_t \vec{L} = J + A \vec{Q} \quad (\text{D.14d})$$

where

$$A = -\hat{\Omega} \quad (\text{D.15a})$$

$$G_\alpha = D_\alpha \vec{V} \quad (\text{D.15b})$$

$$H_\alpha = \partial_\alpha \vec{\Omega} \quad (\text{D.15c})$$

$$K = D_\alpha \vec{F}^\alpha + \vec{f} \quad (\text{D.15d})$$

$$J = D_\alpha \vec{M}^\alpha + \vec{\theta}_\alpha \times \vec{F}^\alpha + \vec{m}. \quad (\text{D.15e})$$

As $A \in \mathfrak{so}(3)$, the short-term propagator for each individual equation in Eq. D.14 can now be found using the $SO(3)$ -integrator defined in Sec. 7.2. The resulting integrator requires far less computation (in particular, it does not require evaluating Eq. D.8), however this comes at the cost of omitting the broader geometric properties of $SE(3)$.

Appendix E

Non-dimensionalisation of the Cosserat rod equations of motion

Here we will derive the non-dimensionalised kinodynamic equations of motion of an underdamped Cosserat rod (given in Eq. 5.120), as well as an overdamped Cosserat rod (given in Eq. 5.125, in the absence of internal frictional and body forces. We set

$$\begin{aligned}\vec{\theta} &= \tilde{\vec{\theta}} \\ \vec{\pi} &= L_0^{-1} \vec{\tilde{\pi}} \\ \vec{V} &= \frac{L_0}{\tau} \vec{\tilde{V}} \\ \vec{\Omega} &= \tau^{-1} \vec{\tilde{\Omega}} \\ \vec{L} &= \frac{\rho_0 L_0^2}{\tau} \vec{\tilde{L}} \\ \vec{F} &= F_0 \vec{\tilde{F}} \\ \vec{M} &= M_0 \vec{\tilde{M}}\end{aligned}$$

where τ , F_0 and M_0 are time, force and moment scales that will be specified later. We also find that $\tilde{P} = \tilde{V}$. As $\tilde{L} = \mathbb{I}\tilde{\Omega}$, we also define the dimensionless moment of inertia $\tilde{\mathbb{I}} = \frac{1}{\rho L_0^2} I$. The resulting equations of motion become

$$D_{\tilde{t}} \vec{\tilde{\theta}} = D_{\tilde{u}} \vec{\tilde{V}} \tag{E.1}$$

$$\partial_{\tilde{t}} \vec{\tilde{\pi}} = D_{\tilde{u}} \vec{\tilde{\Omega}} \tag{E.2}$$

$$\alpha^T D_{\tilde{t}} \vec{\tilde{P}} = D_{\tilde{u}} \vec{\tilde{F}} - \beta^T \vec{\tilde{V}} \tag{E.3}$$

$$\alpha^R D_{\tilde{t}} \vec{\tilde{L}} = D_{\tilde{u}} \vec{\tilde{M}} + \zeta \vec{\tilde{\theta}} \times \vec{\tilde{F}} - \zeta \lambda \vec{\tilde{\Omega}} \tag{E.4}$$

where

$$\begin{aligned}\alpha^T &= \frac{\rho L_0^2 / \tau^2}{F_0} \\ \beta^T &= \frac{(L_0 / \tau) \gamma_T}{F_0 / L_0} \\ \alpha^R &= \frac{\rho L_0^3 / \tau^2}{M_0} \\ \zeta &= \frac{F_0}{M_0 / L_0} \\ \lambda &= \frac{\gamma_R}{\gamma_T L_0^2}\end{aligned}$$

α^T can be seen as the ratio of the characteristic inertial forces compared to the characteristic internal force amplitude F_0 , and we have assumed that the translational and rotation friction coefficients are scalar constants. β^T is the ratio of the characteristic frictional force to the internal force. α^R is the ratio of the characteristic inertial moment to the characteristic moment amplitude M_0 . ζ compares the characteristic force amplitude to the moment. λ compares the translation to the rotational friction.

Without loss of generality we now set $\beta^T = 1$, which means $\tau = \frac{\gamma_T L_0^2}{F_0}$ which is the characteristic damping time-scale for a simple harmonic oscillator. We also set $\zeta = 1$, which ensures that if \vec{F} and \vec{M} are of the same order, then \vec{F} and \vec{M}/L_0 are as well. We also then have that $\alpha^R = \alpha^T = \alpha = \frac{\rho A L_0^2 / \tau^2}{F_0}$. The resulting equations are

$$D_{\tilde{t}} \vec{\theta} = D_{\tilde{u}} \vec{V} \quad (\text{E.5})$$

$$\partial_{\tilde{t}} \vec{\pi} = D_{\tilde{u}} \vec{\Omega} \quad (\text{E.6})$$

$$\alpha D_{\tilde{t}} \vec{V} = D_{\tilde{u}} \vec{F} - \vec{V} \quad (\text{E.7})$$

$$\alpha D_{\tilde{t}} \vec{L} = D_{\tilde{u}} \vec{M} + \vec{\theta} \times \vec{F} - \lambda \vec{\Omega} \quad (\text{E.8})$$

with two tunable parameters α and λ , as well as the dimensionless moment of inertia \tilde{I} .

Let us now assume that we have constitutive laws

$$F_i = g_i \theta_i$$

$$M_i = \epsilon_i \pi_i$$

then in non-dimensionalised form these become

$$\tilde{F}_i = \tilde{g}_i \tilde{\theta}_i$$

$$\tilde{M}_i = \tilde{\epsilon}_i \tilde{\pi}_i$$

where $\tilde{g}_i = g_i/F_0$ and $\tilde{\epsilon}_i = \epsilon_i/(L_0 M_0)$. For overdamped systems we have $\alpha = 0$, such that

$$D_{\tilde{t}} \vec{\theta} = D_{\tilde{u}} \vec{V} \quad (\text{E.9})$$

$$\partial_{\tilde{t}} \vec{\pi} = D_{\tilde{u}} \vec{\tilde{\Omega}} \quad (\text{E.10})$$

$$\begin{aligned} \vec{V} &= D_{\tilde{u}} \vec{F} \\ \lambda \vec{\tilde{\Omega}} &= D_{\tilde{u}} \vec{M} + \vec{\theta} \times \vec{F} \end{aligned} \quad (\text{E.11})$$

Appendix F

Details of the geometric integrator benchmark simulations

F.1 Simulation of the Cosserat rod

We simulated the Cosserat rod with overdamped and underdamped dynamics, specified in Eq. 5.125 and Eq. 5.120 respectively. The non-dimensionalised equations of motion are given in App. E. Henceforth all quantities will be assumed to be non-dimensional, and we will therefore omit the tilde-notation of App. E. For overdamped dynamics we used $\lambda = 1$ and simulated for a duration of $T = 0.1$. For underdamped dynamics we used $\alpha = \lambda = 1$ and $T = 0.1$. The material coordinate took values in the unit interval $u \in [0, 1]$.

To simulate the equations of motion of both systems we discretised space and time as

$$u_i = i\Delta u, \quad i = 0, \dots, N_u \quad (\text{F.1a})$$

$$t_i = i\Delta t, \quad i = 0, \dots, N_t \quad (\text{F.1b})$$

where we used $\Delta u = 1/200$ for all simulations, such that $N_u = 200$. We ran simulations using a range of time-discretisations

$$\Delta t \in \mathcal{T} = \{a10^{-b} : a \in \{1, 1.25, 2, 2.5, 5\} \text{ and } b \in \{4, \dots, 6\}\} \quad (\text{F.2})$$

and for each Δt , we ran simulations where the rod was subject to a randomly configured force and moment, and starting with random initial conditions, as well as random moment of inertia for the underdamped system. We do not consider the effects of internal frictional or body forces.

The angular momentum, and the forces and moments were of the form

$$\vec{L} = \mathbb{I}\vec{\Omega} \quad (\text{F.3a})$$

$$\vec{F} = \mathbf{K}^{(1)}(\vec{\theta} - \vec{\theta}_0) \quad (\text{F.3b})$$

$$\vec{F} = \mathbf{K}^{(2)}\vec{\pi} \quad (\text{F.3c})$$

where $\mathbb{I}, \mathbf{K}^{(1)}, \mathbf{K}^{(2)} \in \mathbb{R}^{3 \times 3}$ were randomly generated each run using the code provided in Listing F.1.

Listing F.1: Generating random stiffness matrices and moment of inertia.

```
import numpy as np
```

```
def random_pos_def_matrix():
    mat = np.random.random((3, 3))
    mat = mat + mat.T
    mat += np.eye(3) * 3
    mat /= np.max(np.linalg.eigvals(mat))
    mat *= 1e-2
    return mat

# Generate random K1 matrix
K1 = random_pos_def_matrix()

# Generate random K2 matrix
K2 = random_pos_def_matrix()

# Generate random moment of inertia
mom_I = random_pos_def_matrix()
```

Initial configurations of the rod $\Phi_0 = (\mathbf{r}_0; E_0)$ were constructed by first generating a random closed center-line $\mathbf{r}_0(u)$ and its corresponding adapted frame, after which the latter was rotated into a deformed material frame $E_0(u)$. The random center-line was generated using a random superposition of sinusoidal functions, using the code provided in Listing F.2.

Listing F.2: Generating random stiffness matrices and moment of inertia.

```
N_random_curve_modes = 3
mu_random_curve = 0
sigma_random_curve = 0.1
L0 = 1
Nu = 200
us_grid = np.linspace(0, 1, Nu, endpoint=False)

def generate_random_periodic_function(dim, us, L, N, mu, sigma):
```

```

fs = np.zeros((3, len(us)))
dfs = np.zeros((3, len(us)))

for i in range(dim):
    fs[i] += np.random.normal(mu, sigma)/2

    for j in range(1, N):
        a, b= np.random.normal(mu, sigma), np.random.normal(mu, sigma)

        fs[i] += a * np.cos((2*np.pi/L)*j*us) / (j*np.pi)
        fs[i] += b * np.sin((2*np.pi/L)*j*us) / (j*np.pi)

        dfs[i] += -(a * (2*np.pi/L)*j / (j*np.pi)) * np.sin((2*np.pi/L)
            *j*us)
        dfs[i] += (b * (2*np.pi/L)*j / (j*np.pi)) * np.cos((2*np.pi/L)*
            j*us)

return fs, dfs

R0, dR0 = generate_random_periodic_function(dim, us_grid, L0,
N_random_curve_modes, mu_random_curve, sigma_random_curve)

```

For a given center-line $\mathbf{r}_0(u)$ we can construct an adapted frame by computing

$$h = |\partial_u \mathbf{r}| \quad (\text{F.4a})$$

$$\mathbf{t} = (\partial_u \mathbf{r})/h \quad (\text{F.4b})$$

$$\mathbf{n} = (t_1 t_3, t_2 t_3, -t_1^2 - t_2^2)^T / \sqrt{t_1^2 + t_3^2 + (t_1^2 + t_2^2)^2} \quad (\text{F.4c})$$

$$\mathbf{b} = \mathbf{t} \times \mathbf{n} \quad (\text{F.4d})$$

which corresponds to a configuration $\Phi^F = (\mathbf{r}; E^F)$, where $E^F = (\mathbf{e}_1^F \ \mathbf{e}_2^F \ \mathbf{e}_3^F)$. The corresponding spatial reconstruction field $X^F = \{\vec{\pi}^F; \vec{\theta}^F\}$ of the adapted frame can be computed as

$$\pi_1^F = (\partial_u \mathbf{n}) \cdot \mathbf{b} \quad (\text{F.5a})$$

$$\pi_2^F = -(\partial_u \mathbf{t}) \cdot \mathbf{b} \quad (\text{F.5b})$$

$$\pi_3^F = (\partial_u \mathbf{t}) \cdot \mathbf{n} \quad (\text{F.5c})$$

$$\theta_1^F = h \quad (\text{F.5d})$$

$$\theta_2^F = \theta_3^F = 0. \quad (\text{F.5e})$$

Now, given some rotation matrix $R(u)$ defied over the length of the rod, we construct the

initial configuration of the rod as

$$\hat{\pi}_0 = (\partial_u R)R^T + R\hat{\pi}^F R^T \quad (\text{F.6a})$$

$$\vec{\theta}_0 = R\vec{\theta}^F. \quad (\text{F.6b})$$

The form of these equations are such that the configuration $\Phi_0 = (\mathbf{r}; E_0)$, which are the initial conditions used in the simulations, shares the same center-line with Φ^F , where $E_0 = RE^F$. Equation F.6 can be found by substituting $E_0 = RE$ into Eq. 5.42b. Derivatives of the spatial reconstruction fields $X(t_i, u_i)$ and all other quantities were computed using Fourier transforms, following the numerical algorithms provided in [67].

The deformational rotation R was constructed by generating random Euler angle-fields $\rho(u)$, $\phi(u)$ and $\psi(u)$ along $u \in [0, T]$, and then letting $R = R(\rho, \phi, \psi)$, where $R(\rho, \phi, \psi)$ is the mapping from Euler angles to rotation matrices. The Listing F.3 shows the code that generates random deformations R .

Listing F.3: Generating random deformational rotation field.

```
N_R_modes = 3
mu_R = 0.1
sigma_R = 0.01

def eul2rot(rho, phi, psi) :
    R = np.zeros((3, 3, rho.shape[-1]))
    R[0,0] = np.cos(phi)*np.cos(psi)
    R[0,1] = np.sin(rho)*np.sin(phi)*np.cos(psi) - np.sin(psi)*np.cos(rho)
    R[0,2] = np.sin(phi)*np.cos(rho)*np.cos(psi) + np.sin(rho)*np.sin(psi)

    R[1,0] = np.sin(psi)*np.cos(phi)
    R[1,1] = np.sin(rho)*np.sin(phi)*np.sin(psi) + np.cos(rho)*np.cos(psi)
    R[1,2] = np.sin(phi)*np.sin(psi)*np.cos(rho) - np.sin(alprhoha)*np.cos(psi)

    R[2,0] = -np.sin(phi)
    R[2,1] = np.sin(rho)*np.cos(phi)
    R[2,2] = np.cos(rho)*np.cos(phi)

    return R

us_grid = np.linspace(0, 1, Nu, endpoint=False)
R_rho = generate_random_periodic_function(1, us, L0, N_rot_modes, mu_rot,
    sigma_rot)[0]
R_phi = generate_random_periodic_function(1, us, L0, N_rot_modes, mu_rot,
    sigma_rot)[0]
R_psi = generate_random_periodic_function(1, us, L0, N_rot_modes, mu_rot,
    sigma_rot)[0]
```

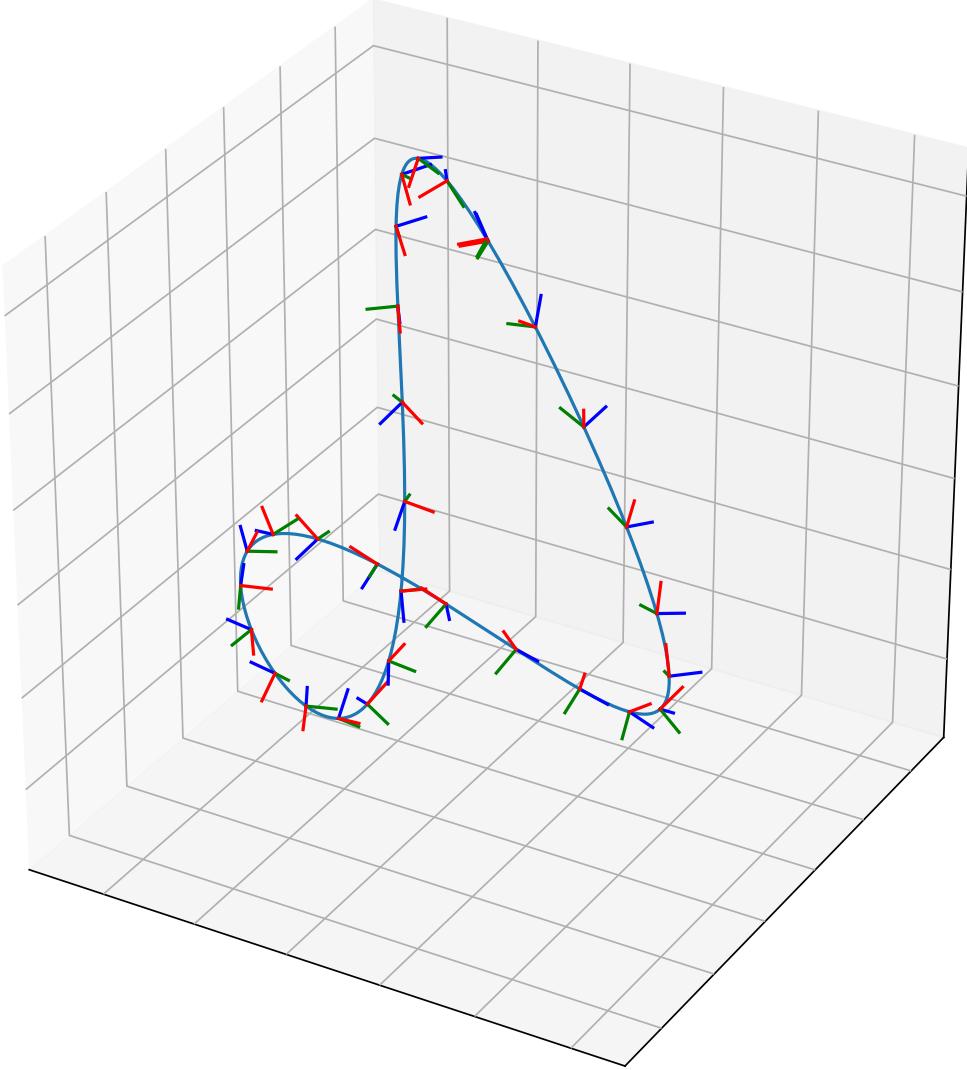


Figure F.1: An example of a random initial configuration of the Cosserat rod generated by the outlined procedure. The blue line is the center-line $\mathbf{r}(u)$ and the red, green and blue lines attached to the center-line are $\mathbf{e}_1(u)$, $\mathbf{e}_2(u)$ and $\mathbf{e}_3(u)$ respectively.

```
R = eul2rot(R_rho, R_phi, R_psi)
```

This completes the description of the procedures used to generate the initial configurations of the Cosserat rod. See Fig. F.1 for an example of a random initial Cosserat configuration. For a given random configuration, which we will refer to via the index ℓ , we ran simulations using the Forward-Euler (FEI), the $SE(3)$ -integrator (SEI) and the simplified $SE(3)$ (SSEI) integrators for each $\Delta t \in \mathcal{T}$. Let $X^{I,\Delta t,\ell}$ be the result of a simulation using a given algorithm $I \in \{\text{FEI}, \text{SEI}, \text{SSEI}\}$, time-step Δt and initial configuration ℓ . Using the algorithm described in App. C.1 we can then reconstruct the corresponding

spatio-temporal configuration $\Phi_\ell^{I,\Delta t,\ell} = (\mathbf{r}_\ell^{I,\Delta t,\ell}; E_\ell^{I,\Delta t,\ell})$. We now define

$$\text{err}_{\text{close}}^{I,\Delta t,\ell} = |\mathbf{r}^{I,\Delta t,\ell}(T, 1) - \mathbf{r}^{I,\Delta t,\ell}(T, 0)| + \sum_i |\mathbf{e}_i^{I,\Delta t,\ell}(T, 1) - \mathbf{e}_i^{I,\Delta t,\ell}(T, 0)| \quad (\text{F.7a})$$

which measures the failure of the Cosserat rod to close. That is, $\text{err}_{\text{close}}^{I,\Delta t,\ell}$ is the magnitude of the discontinuity of $\Phi_\ell^{I,\Delta t,\ell}$ at $u = 1$. As the initial configuration is a closed loop, we have $\mathbf{r}^{I,\Delta t,\ell}(0, 0) = \mathbf{r}^{I,\Delta t,\ell}(0, 1)$ for all simulations. Simulation errors will however in general cause the configuration of the Cosserat rod in the numerics to no longer close. Note that it is not sufficient for $X^{I,\Delta t,\ell}(t, u)$ to be a smooth periodic function of u in order for $\Phi_\ell^{I,\Delta t,\ell}$ to be smooth a smooth function of u .

For both the underdamped and overdamped Cosserat rod, and for each I and Δt we ran $N_{\text{sim}} = 50$ simulations with random initial conditions $\ell = 1, \dots, N_{\text{sim}}$. We then computed the sample means and variances of the error

$$\langle \text{err}_{\text{close}}^{I,\Delta t} \rangle = \frac{1}{N_{\text{sim}}} \sum_{\ell=1}^{N_{\text{sim}}} \text{err}_{\text{close}}^{I,\Delta t,\ell} \quad (\text{F.8a})$$

$$\text{Var}(\text{err}_{\text{close}}^{I,\Delta t}) = \frac{1}{N_{\text{sim}} - 1} \sum_{\ell=1}^{N_{\text{sim}}} (\text{err}_{\text{close}}^{I,\Delta t,\ell} - \langle \text{err}_{\text{close}}^{I,\Delta t} \rangle)^2. \quad (\text{F.8b})$$

F.2 Simulation of the Cosserat surface

We simulated the kinematic equations of motion of a Cosserat surface, which were given in Eq. 6.92. We discretised space and time as

$$u_i = 1 - \cos\left(\frac{i\pi}{N}\right), \quad i = 0, 1, \dots, N_u \quad (\text{F.9a})$$

$$v_i = 1 - \cos\left(\frac{i\pi}{N}\right), \quad i = 0, 1, \dots, N_v \quad (\text{F.9b})$$

$$t_i = i\Delta t \quad (\text{F.9c})$$

where Eq. F.9a are known as the *Chebyshev points*.¹ We differentiated functions on Chebyshev grids using *spectral differentiation*, see [67] for details. For all numerics we used $N_u = N_v = 20$.

To test our suite of integrators we evaluated their short-time propagators using a range of time-discretisations

$$\Delta t \in \mathcal{T} = \{a10^{-b} : a \in \{1, 1.25, 2, 2.5, 5\} \text{ and } b \in \{4, \dots, 7\}\}. \quad (\text{F.10})$$

Every short-term propagator was evaluated on the same initial condition $X_{0,\alpha}(u, v) =$

¹Eq. F.9a is however in a non-standard form, as we have ordered the points in ascending order.

$\{\vec{\theta}_{0,\alpha}, \vec{\pi}_{0,\alpha}\}$, $\alpha = u, v$, where

$$(\theta_{0,u})_2(i\Delta u, j\Delta u) = 1 \quad (\text{F.11a})$$

$$(\theta_{0,u})_1(i\Delta u, j\Delta u) = (\theta_{0,u})_3(i\Delta u, j\Delta u) = 0 \quad (\text{F.11b})$$

$$(\theta_{0,v})_3(i\Delta u, j\Delta u) = 1 \quad (\text{F.11c})$$

$$(\theta_{0,v})_1(i\Delta u, j\Delta u) = (\theta_{0,v})_2(i\Delta u, j\Delta u) = 0 \quad (\text{F.11d})$$

$$\vec{\pi}_{0,u}(i\Delta u, j\Delta u) = \vec{\pi}_{0,v}(i\Delta u, j\Delta u) = 0 \quad (\text{F.11e})$$

for all $i = 0, \dots, N_u$, $j = 0, \dots, N_v$. The corresponding spatio-temporal configuration $\Phi_0 = (\mathbf{r}, E)$ is a flat sheet $\mathbf{r}(i\Delta u, j\Delta u) = (0, i\Delta u, j\Delta v)^T$ with a constant perpendicular material frame. The system is subject to generalised velocity field $N_0 = \{\vec{V}_0, \vec{\Omega}_0\}$ which we generate as a random superposition of sinusoidal surfaces, which can be reproduced using Listing F.4.

Listing F.4: Setting up initial conditions and generating random velocity fields.

```
# Parameters

Lu0 = 1
Lv0 = 1

Nu = 20
Nv = 20

N_rand_V_u = 5
N_rand_V_v = 5

N_rand_Omg_u = 5
N_rand_Omg_v = 5

rand_V_var = 1e0
rand_Omg_var = 1e0

# Initial conditions

thu0 = np.zeros((3, Nu, Nv))
thv0 = np.zeros((3, Nu, Nv))
piu0 = np.zeros((3, Nu, Nv))
piv0 = np.zeros((3, Nu, Nv))

thu0[1] = 1
thv0[2] = 1

piu0[1] = 0
```

```

piv0[0] = 0

# Construct material mesh grid

def get_grid_1D(Mm, L):
    N = Mm - 1
    x = np.cos(np.pi*np.arange(0,N+1)/N)
    x = L*(1-x)/2
    return x

us = get_grid_1D(Nu, Lu0)
vs = get_grid_1D(Nv, Lv0)
grid_U, grid_V = np.meshgrid(us, vs, indexing='ij')

# Define random velocity and angular velocity

rand_V_coeffs = np.zeros((3, N_rand_V_u, N_rand_V_v))
rand_Omg_coeffs = np.zeros((3, N_rand_Omg_u, N_rand_Omg_v))

for di in range(dim):
    for i in range(N_rand_V_u):
        for j in range(N_rand_V_v):
            rand_V_coeffs[di,i,j] = rand_V_var * np.random.normal()
            rand_Omg_coeffs[di,i,j] = rand_Omg_var * np.random.normal()

# Compute velocity field

V_field = np.zeros((3, Nu, Nv))
for di in range(dim):
    for i in range(N_rand_V_u):
        for j in range(N_rand_V_v):
            V_field[di] += rand_V_coeffs[di,i,j] * np.sin((i - 0.5)*np.pi*
                grid_U/Lu0) * np.sin((j - 0.5)*np.pi*grid_V/Lv0) / ( (i-0.5)
                * (j - 0.5) * np.pi**2 )

# Compute angular velocity field

Omg_field = np.zeros((3, Nu, Nv))
for di in range(dim):
    for i in range(N_rand_Omg_u):
        for j in range(N_rand_Omg_v):
            Omg_field[di] += rand_Omg_coeffs[di,i,j] * np.sin((i - 0.5)*np.
                pi*grid_U/Lu0) * np.sin((j - 0.5)*np.pi*grid_V/Lv0) / ( (i
                -0.5) * (j - 0.5) * np.pi**2 )

```

Let $I \in \{\text{FEI}, \text{SEI}, \text{SSEI}\}$ refer to the Forward-Euler, $SE(3)$ - and simplified $SE(3)$ -

propagators respectively, and let $\ell = 1, \dots, N_{\text{sim}}$ be an index over the set of randomly generated velocity fields. In our results we used $N_{\text{sim}} = 50$. Let $X_{\Delta t, \alpha}^{I, \ell}$ be the result of computing the short-term propagator using integrator I , with time-step Δt and configuration ℓ . For each I , Δt and ℓ , we computed the supremum of the residual integrability error Δ_{uv}^{int} , given in Eq. 6.9 over the material base space. In other words we compute

$$\text{err}_{\text{int}}^{I, \Delta t, \ell} = \sup_{i=1, \dots, N_u, j=1, \dots, N_v} \left| \partial_v X_{\Delta t, u}^{I, \ell}(i \Delta u, j \Delta v) - \mathcal{D}_u X_{\Delta t, v}^{I, \ell}(i \Delta u, j \Delta v) \right| \quad (\text{F.12})$$

where the derivatives were computed on the Chebyshev grid using spectral differentiation. We also compute the mean and variance of the integrability error over the set of configurations ℓ

$$\langle \text{err}_{\text{int}}^{I, \Delta t} \rangle = \frac{1}{N_{\text{sim}}} \sum_{\ell=1}^{N_{\text{sim}}} \text{err}_{\text{int}}^{I, \Delta t, \ell} \quad (\text{F.13a})$$

$$\text{Var}(\text{err}_{\text{int}}^{I, \Delta t}) = \frac{1}{N_{\text{sim}} - 1} \sum_{\ell=1}^{N_{\text{sim}}} (\text{err}_{\text{int}}^{I, \Delta t, \ell} - \langle \text{err}_{\text{ref}}^{I, \Delta t} \rangle)^2. \quad (\text{F.13b})$$