Multiscale Geometric Integration of Deterministic and Stochastic Systems

Thesis by
Molei Tao

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2011

(Defended May 16, 2011)

For my parents,

and

In memory of Professor Jerrold Eldon Marsden.

And if I were allowed to be greedy —

May the entire human race unite,

despite race, nationality, social class, or gender.

We are fragile and small, after all.

Acknowledgements

This thesis would not be possible without the support of many people, and I am deeply grateful to all of them. Since my arrival at Caltech, I had the great honor to be supervised by Houman Owhadi and Jerrold E. Marsden. No matter what subject I wish to learn, whether it is about analysis, probability, or geometry, they are always able to give me an incredible amount of help. The recent loss of Dr. Marsden immensely saddened my life, but hoping that one day I can proudly announce myself as one of his legacies, I will firmly carry on our work. Prof. Owhadi continues to educate me on all aspects as always: besides training my analytical abilities, for five years he has kept on urging me to form a clear big picture and to pursue interesting and important directions, providing me warm recommendations and various presentation opportunities, introducing me to exciting problems, and preparing me to communicate mathematics professionally. His brilliance, deep insight, optimism, and enthusiasm provided the perfect role model for me. I cannot think of a way to return the favor, except to strive to become a better researcher.

Another pair of people to whom I could never return their immeasurable care are my parents. There are no words that can summarize the things that we went through together. I am not a celebrity, but mom and dad, you are always the perfect fans. Also, I would like to thank my girlfriend, Gongjie Li, whose unconditional support makes me feel at home. I hope I can satisfy my responsibilities and deliver smiles to them.

I would also like to express my sincere gratitude to other members of my thesis committee: Richard Murray, Mathieu Desbrun, Thomas Hou, and Michael Ortiz. I really appreciate their kind help, from which I benefit in many aspects, including the content of my thesis work, the development of my career, etc. I also received academic help from other faculty, including Richard Y. Tsai (stimulating discussion on stiff ODEs and kind help on career development), James Beck (valuable discussion and detailed suggestions on temperature and friction accelerated sampling), and Eric Vanden-Eijnden (insightful discussion on the String method), to

whom I am greatly thankful.

Moreover, I am delighted to have many brilliant collaborators and friends. Just to name a few: Sina Ober-Blöbaum, Julia Greer, Yun Jing, Mulin Cheng, Nawaf Bou-Rabee, Wang Soon Koon, and Tomohiro Yanao. At the same time, I wish to acknowledge an incomplete list of my colleague and friends: Andrew Benson, Vanessa Jonsson, Matanya Horowitz, Shaun Maguire, Dennice Gayme, Konstantin Zuev, Guo Luo, Alex Gittens, Na Li, Andrew Lamperski, Henry Jacobs, Andrea Censi, Francisco Zabala, Eric Wolff, Tichakorn (Nok) Wongpiromsarn, Paul Skerritt, Sawyer Fuller, Dionysios Barmpoutis, Joris Vankerschaver, Elisa Franco, Gentian Buzi, Nader Motee, Hamed Hamze Bajgiran, David Pekarek, and Ari Stern. I also wish to thank many other friends who made my graduate studies memorable and educational, including: Shuo Han, Chengshan Zhou, Hongchao Zhou, Maolin Ci, Ling Shi, Hongjin Tan, Xin Heng, Gerardo Cruz, Yuan Zhong, Jie Wu, Liming Wang, Yue Shen, Yizhou Liu, Na Li, Huan Yang, Ting Hong, Jie Cheng, Ying Wang, Fei Wang, Sijia Dong, Gregory Kimball, Andrej Svorencik, Tian Lan, Le Kuai, Desmond Cai, Prabha Mandayam, Xi Dong, Ding Weng, Zhouyuan Zhu, Yu Zhou, Jialin Zhang, Shunan Zhang, Chengjie Wu, Zhongkai Liu, Yan Jiang, Miaomiao Fan, Yuan Zhang, Xiaowei Zhang, Yuan Zhong, Qin Sun, Kui Wang, Xiaochuan Zhang, Lin Wang, Yifei Ding, and Jiaojie Luan.

Among the administration at Caltech, I wish to especially thank Sydney Garstang, Icy Ma, and Wendy McKay for their kindness. I also feel extremely fortunate for having been able to attend Caltech, Control & Dynamical Systems, and Computing & Mathematical Sciences, where flexibility, resource abundance, a motivated atmosphere, and an open environment promote collaboration, concentration, excellency, and creativity.

Multiscale Geometric Integration of Deterministic and Stochastic Systems

by

Molei Tao

Abstract

In order to accelerate computations and improve long time accuracy of numerical simulations, this thesis develops multiscale geometric integrators.

For general multiscale stiff ODEs, SDEs, and PDEs, FLow AVeraging integratORs (FLAVORs) have been proposed for the coarse time-stepping without any identification of the slow or the fast variables. In the special case of deterministic and stochastic mechanical systems, symplectic, multisymplectic, and quasi-symplectic multiscale integrators are easily obtained using this strategy.

For highly oscillatory mechanical systems (with quasi-quadratic stiff potentials and possibly high-dimensional), a specialized symplectic method has been devised to provide improved efficiency and accuracy. This method is based on the introduction of two highly nontrivial matrix exponentiation algorithms, which are generic, efficient, and symplectic (if the exact exponential is symplectic).

For multiscale systems with Dirac-distributed fast processes, a family of symplectic, linearly-implicit and stable integrators has been designed for coarse step simulations. An application is the fast and accurate integration of constrained dynamics.

In addition, if one cares about statistical properties of an ensemble of trajectories, but not the numerical accuracy of a single trajectory, we suggest tuning friction and annealing temperature in a Langevin process to accelerate its convergence.

Other works include variational integration of circuits, efficient simulation of a nonlinear wave, and finding optimal transition pathways in stochastic dynamical systems (with a demonstration of mass effects in molecular dynamics).

Contents

Chapter 1

Introduction

When I was shooting birds at pigs on my mobile phone (amazingly, this is a very popular game in 2010 and 2011) waiting for a take-away order, I was suddenly struck by the thought that my phone exceeds the sum of all computational powers employed in 1970 to send humans to the moon. How can I be part of this restless development of technology? I may not be able to immediately contribute to advances in hardware, but I could manipulate equations. And by manipulating equations, serious scientific computations can be made much cheaper. Soon, I will be able to do something more on my phone than slingshotting birds.

1.1 Necessity of numerical integration

Differential equations of various types, including ordinary differential equations (ODEs), stochastic differential equations (SDEs), partial differential equations (PDEs), and differential algebraic equations (DAEs), are mathematical tools for describing changes in a system. Therefore, their importance in natural sciences, engineering and social sciences is needless to mention. In addition, more and more modern entertainment, ranging from the bird-shooting video game to 3D blockbusters, are based on modeling using differential equations.

Solving differential equations, however, is not always an easy task. Firstly, nonlinearity in the equation oftentimes eliminates the possibility of obtaining a closed-form analytical solution. For example, the simple ODE $\ddot{\theta} = \sin\theta$ that mod-

els a pendulum cannot be solved exactly. In addition, even if analytical expressions are available, they are not necessarily easy to manipulate. For instance, the simple ODE $m\ddot{q} + c\dot{q} + kq = 0$ that models a damped harmonic oscillator has a closed-form solution, but it is a long expression and difficult to work with. Moreover, the mathematical investigation of the existence and/or uniqueness of a solution is highly nontrivial in many cases. Also, there are different senses in which a solution could solve the system. For instance, an ODE with an initial condition may have a solution that exists only until a finite time, or it could admit a class of solutions [228]; an SDE with an initial condition might have different solution in the sense of Itō or Stratonovich [219]; a PDE could have no solution in the strong sense, yet still admit a single or even multiple weak solutions [99]; a DAE with an initial condition may have zero or many solutions, just due to its ODE component, which could be further complicated by the additional algebraic constraints.

Numerical integrations partially solve these issues. With the aid of modern computers, many nonlinear differential equations can be numerically integrated (we oftentimes use the word 'integrate' to mean 'solve' in the context of differential equations). In addition, the sense in which these equations are solved is often assumed by the integrator; for instance, finite difference method assumes the existence of a smooth strong solution, while finite element method is based on a weak formulation.

1.2 Limitations of numerical integration

1.2.1 Computational costs in terms of both time and memory

As a long-standing challenge, the problem of numerical integration is far from being completely solved. A first difficulty is that a traditional numerical integration of a complex system consumes a significant amount of time and memory.

Multiscale systems are a particular example. Specifically, if the system exhibits dynamics on different timescales, for instance when its governing equations contain a stiff parameter, then traditional integrators require resolving the fastest

timescale so that the correct slow timescale can be obtained. This, obviously, is not computationally efficient, and an integration that uses a coarse timestep, which corresponds to the slow timescale, is often desired.

For instance, it could take a traditional integrator $\sim 10^{12}$ integration steps to simulate the folding of a protein. A protein usually takes milliseconds to fold, but fast components of its dynamics, such as bond oscillations, happen at the timescale of picoseconds [26]. Since these components contribute to the global dynamics in a nontrivial way, traditional integration requires them to be resolved. This, however, was not practically feasible until the recent development of a specialized super computer [253], which nevertheless still spent months on such a computation. Despite being computationally expensive, such numerical simulations are vital to scientific studies because they are still much cheaper than in vivo or in vitro experiments, and they provide microscopic details that are beyond the accuracy of contemporary experimental measurements.

On top of the difficulty in bridging different timescales, the dimensionality of the system also incurs computational expenses. Consider for example the evolution of the universe, whose simulation is of great cosmological interest. In the famous Millennium Simulation [263], researchers used an N-body simulation with $N \approx 10^{10}$ to reproduce the history of our universe; the price for their fruitful investigation of cosmology is a one-month simulation (based on a classical algorithm of symplectic leap-frog), 512 processors, and 700 GB memory, which is beyond the computational capacity of most applied math labs in the year of 2011.

Notice that the N-body simulation of universe evolution not only involves a large number of variables but also exhibits dynamics over multiple timescales. Unlike the protein dynamics for which the presence of stiff parameters induces a separation of timescales, the origin of multiple timescales in the N-body model is due to nonlinearities in the corresponding dynamical system. In fact, nonlinearities also manifest in protein models (mainly contained in noncovalent forces). As a consequence, the slow dynamics is further split into a slow scale and a slower scale, resulting in at least three timescales in the protein dynamics.

Of course, in the case of PDE, the system could exhibit not only different timescales but also different spatial scales.

1.2.2 Inaccuracy of long time simulation

The second challenge in numerical integration is the accumulation of numerical error with increased integration time. The textbook error bounds are of the form $C\exp(CT)h^p$ (see for instance [129]), where T is the total simulation time, h is the size of the integration timestep, C is a positive constant that depends on the derivative of the vector field, and p is another positive constant indicating the order of convergence. This means that for a fixed T, the integration can be made accurate by choosing h small enough, but no matter how small h is, the error may blow up exponentially in a long time simulation. This issue worsens in multiscale systems, because, by the time the slow timescale is reached, errors from the fast timescale will have already accumulated intensively. This could be illustrated by an example of a stiff system: indicate the stiff parameter as ϵ^{-1}, then in the worst case the above error bound is written as

$$\epsilon^{-1}C\exp(\epsilon^{-1}CT)h^p \tag{1.1}$$

where the constant C in the classical error bound is replaced by $\epsilon^{-1}C$ due to the stiffness contained in the vector field. Consequently, the error blows up (as $\epsilon \to 0$) at $T = \mathcal{O}(1)$.

Interestingly, this illustrates that rapid advances in computer hardware alone cannot relieve the concern on computational efficiency of numerical integrations. In fact, we require algorithmic breakthroughs regardless of the availability of computational power. The reason is the following: with a fast enough computer, we can choose a small integration step to simulate a complex system with high accuracy till time $\mathcal{O}(1)$; however, no matter how small this step is, the integration will not be accurate at an arbitrary time T due to the exponential growth with T in (1.1), unless sophisticated methods specifically designed for long time simulations

are proposed.

1.3 Examples of state-of-the-art numerical approaches

This section discusses an incomplete list of contemporary efforts towards solving the problems of multiscale integration and long time simulation (without including this thesis' contribution). Details of the methods and rigorous definitions of terminologies will not be described here, but relevant information could be found in later chapters of this thesis.

1.3.1 Structure preserving methods

One way to improve long time numerical integrations is to utilize structures (many of which are geometric) in the system of interest. *'The subject of geometric numerical integration deals with numerical integrators that preserve geometric properties of the flow of a differential equation, and it explains how structure preservation leads to an improved long-time behavior'* [131].

Mechanical systems: Mechanical systems conserve energy and momentum maps (such as linear momentum and angular momentum; a slightly more modern example is the charge conservation due to a $U(1)$ symmetry in quantum field theory [264]), and their solution flows preserve an underlying geometric structure of symplecticity (multisymplecticity in the case of PDE), which intuitively means that any infinitesimal volume in the phase space will be preserved. All these conservations are consequences of an underlying variational structure in the system (details can be found, for instance, in [3, 194]).

Structure preserving numerical methods for Hamiltonian systems have been developed in the framework of geometric numerical integration [128, 179], and various structures have been addressed by different approaches. For instance, symmetric methods are based on the reversibility of their updating maps, and thus have good long time performance [128]; energy-momentum methods enforce

the conservation of momentum by their updating rules [254]; Lie-group integrators ensure that the numerical solution stays in a desired Lie-group by updating rules obtained from a geometric computation [154].

What is worth special emphasis is the family of variational integrators, for it might be the method that preserves the most structures so far. Variational integration theory derives integrators for mechanical systems from discrete variational principles that correspond to discrete mechanics [192]. Therefore, a variational integrator naturally preserves a discrete symplectic form, obeys a discrete Noether's theorem (and therefore preserves discrete momentum maps), and nearly conserves the energy in the system because it in fact yields the exact solution of a nearby mechanical system (due to backward error analysis). Variational integrators fall in a larger category of symplectic integrators (see [245] for a review on symplectic integrators). On the converse, symplectic integrators are at least locally variational [128], and therefore they usually have similar preservation properties as variational integrators.

The preserved structures in symplectic integrators certainly help long time numerical integrations. An intuitive illustration is, the (near) preservation of energy in a harmonic oscillator rules out the possibility of any exponential growth in error, because otherwise the energy will not remain bounded. In fact, it has been shown that symplectic integrators for integrable systems have an error bound that is linearly growing with the integration time [131]. A well-known numerical observation is, no matter how small a time step is used, the oscillation amplitude of a harmonic oscillator integrated by non-symplectic Forward Euler/Backward Euler will increase/decrease unboundedly, whereas that given by Variational Euler (also known as symplectic Euler) will be oscillatory with a variance controlled by the step length.

Other notable properties of variational integrators include: *'Variational integrators can readily incorporate holonomic constraints (via, e.g., Lagrange multipliers) and nonconservative effects (via, e.g., their virtual work)'* (quoted from [40] with references [295, 192]). In addition, variational integrators can handle

nonholomonic mechanical systems and degenerate Lagrangian systems, because these systems can be formulated in the context of implicit Lagrangian systems (associated with a Dirac structure) [302], which have a variational structure based on the Hamiltonian-Pontryagin-d'Alembert principle [303]. Furthermore, statistical properties of the dynamics, such as Poincaré sections, are well preserved by variational integrators even with large time steps [39].

Stochastic systems: For a system based on SDEs with geometric ergodicity, temporal averaging of its long time behavior converges to the spatial average with respect to its corresponding ergodic measure. A strategy has been proposed to provide numerical approximations that satisfy an analogous (discrete) geometric ergodicity [198], which according to the authors is *'the first step in an analysis of the convergence of invariant measures of discretizations to those of the SDE itself'* following *'the pioneer work'* of [272].

For the special case of stochastic mechanical systems, *'since the foundational work of Bismut [33], the field of stochastic geometric mechanics is emerging in response to the demand for tools to analyze continuous and discrete mechanical systems with uncertainty'* [40]. For instance, an incomplete list of integrators for Langevin equations, which model mechanical systems under dissipation and perturbation by external noises, include [255, 136, 290, 69, 203, 204, 206, 175, 188, 40, 41, 42]. One interesting result is that the composition of the one-step update of a variational integrator and an Ornstein-Uhlenbeck process will produce a good numerical solution — good in the sense that the numerical approximation converges to an ergodic measure that is close in total variation norm to the Boltzmann-Gibbs ergodic measure associated to the exact solution [41]. Indeed, in the case of Langevin, the long time accuracy in terms of statistics is a natural stochastic extension of the preservation of a near-by energy function. While the preservation of energy can be schematically viewed as that the solution stays in a constant energy submanifold with probability one, which can be numerically approximated by a symplectic integrator, the numerical preservation of a near-by

Boltzmann-Gibbs measure has its root in the quasi-symplecticity [205] and/or the conformal symplecticity [202] of the corresponding integrator.

One interesting note is that there are deterministic chaotic systems that are ergodic (e.g., Lorenz attractor [185]), and a natural thought would be to devise an integrator that nonintrusively produces numerical approximations that are ergodic with respect to a measure close to the exact ergodic measure. I am aware of few existing approaches achieving this possibility.

Conservation law PDEs: PDEs of this type satisfy the Rankine-Hugoniot conditions [99], which provide an important characterization of shock propagations. Finite volume methods used on conservation law PDEs, for instance, obey discrete conservation laws, satisfy analogous Rankine-Hugoniot conditions, and therefore work for shock capturing [182]. Moreover, if carefully designed (for instance, Godunov's method [120]), a conservative numerical scheme is able to pick the solution that satisfies the correct entropy condition among a family of weak solutions. The satisfaction of the Rankine-Hugoniot identity and the entropy inequalities certainly benefits long time simulations.

Hamiltonian PDEs: Hamiltonian PDEs are a special class of infinite dimensional mechanical systems. Naturally, structures in mechanical systems, such as the conservations of momentum maps, which in the continuous setting are guaranteed by Noether's theorem from symmetries, will be important to long time simulations. A brief review of Hamiltonian PDEs, as well as a numerical recipe of multisymplectic integrators which satisfy a discrete Noether's theorem, can be found in Section 3.2.1. Hamiltonian PDEs are not to be confused with Hamilton-Jacobi PDEs (reviewed in, for instance, [99]).

Oftentimes, structure preserving ODE integrators such as those described above are called geometric integrators, because they preserve various geometric properties. We will call structure preserving SDE/PDE solvers geometric integrators as well.

1.3.2 Multiscale methods

'Dynamical systems with multiple time scales pose a major problem in simulations because the small time steps required for stable integration of the fast motions lead to large numbers of time steps required for the observation of slow degrees of freedom' [286]. Regarding the numerical integration of multiscale systems with coarse steps, a large variety of methods are applicable to different systems. A large portion of them have been devoted to stiff systems, in which the presence of a large parameter gives rise to a separation of timescales.

Stiff ODEs and SDEs: Traditionally, stiff dynamical systems based on ODEs and SDEs have been separated into two classes with distinct integrators: stiff systems with fast transients and stiff systems with rapid oscillations [14, 85, 242]. The former have been solved using implicit schemes [112, 79, 128, 130, 304], Chebyshev methods [177, 1] or the projective integrator approach [111]. The latter have been solved using filtering techniques [110, 168, 246] or Poincaré map techniques [113, 229]. We also refer to methods based on highly oscillatory quadrature [74, 152, 151], an area that has undergone significant developments in the last few years [153].

When slow variables can be identified, different types of fast processes can be handled in a unified framework, and asymptotically their effective contribution to the slow process could be described analytically by an averaging theorem (see for instance [258, 226, 224, 225, 239]). Two classes of numerical methods have been built on this observation: The equation-free method [165, 164, 15] and the Heterogeneous Multiscale Method (HMM) [86, 97, 85, 13, 87] (as well as its variant, the seamless method [89]).

We further review continuous and numerical treatments of SDE asymptotic problems in Section 2.3.1.

Stiff PDEs: A more difficult case is stiff PDEs. If the system exhibits fast transients (i.e., fast variables convergent towards a Dirac point distribution, or

equivalently (using the terminology in [102]), 'asymptotically stable'), asymptotic preserving schemes [102] based on implicit methods allow for simulations with large time steps. We also refer to [159, 211] for multiscale transport equations and hyperbolic systems of conservation laws with stiff diffusive relaxation.

For cases in which the slow process can be identified, we refer to [86] for a review that includes various applications; also, for a discrete KdV-Burgers' type equation with well-identified fast and slow variables, a coarse time-stepping of the system can be achieved via the equation-free approach [19].

PDEs in homogenization theory: Multi-scale PDEs can be divided into two (possibly overlapping) categories: PDEs with large (or stiff) coefficients and PDEs with highly oscillating or rough coefficients. This thesis only considers the first category.

The second type is the subject of asymptotic and numerical homogenization theory. An example of equations in this category is the following elliptic PDE

$$-\operatorname{div}(a\nabla u) = f \tag{1.2}$$

with the coefficient $a(\cdot)$ being highly oscillatory, random (stationary and ergodic), or rough. Since homogenization is a profound field and it is not the scope of this thesis, we just refer to an incomplete list of continuous and numerical treatments in [5, 23, 27, 70, 157, 223, 167, 167, 43, 116, 117, 210, 209, 261, 262, 296, 44, 146, 147, 68, 86, 96, 119, 21, 22, 6, 20, 220, 29, 221, 34, 35, 86, 96, 30, 92, 216, 281, 135, 54, 11, 12].

One explanation to why homogenization works is based on the compactness of the solution space (i.e., when f spans the unit ball of L^2, u spans a (strongly) compact subset of H^1, which can be approximated in H^1-norm by a finite-dimensional space). This is different from the approaches introduced in this thesis, which are instead based on the separation between slow and fast timescales in systems of the first category (in many cases we also make use of the local ergodicity of the fast timescale).

Other multiscale systems: Another complicated case of multiscale systems is when the equations do not explicitly contain a stiff parameter but there is a separation of timescales, which is usually due to the nonlinearity of the equation and/or initial/boundary conditions. If the fast process in such a system is transient, implicit schemes should work for both ODEs and PDEs. If the slow variable can be explicitly identified, there are examples to which HMM applies [86]; if the slow variable is not identified, but is however known to be a linear function of the original coordinate, there is also an example in which the seamless method works [89].

In addition, [148] provides an example to represent a continuum-scaled system of 3D incompressible Navier-Stokes PDE in two scales by adopting a cut off in the Fourier domain and then treating the problem using a homogenization approach.

1.3.3 Structure preserving multiscale methods

Efforts have been made to combine structure preserving integrators and multiscale methods together (see Equation (1.1) for a motivation). The following is an incomplete list of examples:

Highly oscillatory mechanical systems: For Hamiltonian systems with stiff potentials that are quadratic (such systems are often called highly oscillatory mechanical systems), there are at least two types of numerical integrators, which we will briefly discuss in the following presentation (we also refer to [71] for a recent review). The first type does not rely on an identification of fast or slow variables, and includes, for instance, the following:

Impulse methods [297, 124, 286] are symplectic integrators that admit a uniform error bound on the positions when the potential energy is a sum of an arbitrary soft potential and a quadratic stiff potential [276], and this uniform bound justifies the use of a large integration timestep. In their abstract form, impulse methods are not limited to quadratic stiff potentials; however, their practical implementation requires an approximation of the flow associated with the stiff po-

tential, which in most non-quadratic cases could only be obtained via a small-step integration, and is therefore computationally expensive.

Impulse methods have been mollified [109, 240] to gain extra stability and accuracy while the mollified method remains symplectic. Both mollified impulse methods and Gautschi-type integrators [143] (reversible but not necessarily symplectic any more) can be shown to be members of the exponential integrator family [123]. It has been proved that these methods allow large-time-stepping of Hamiltonian systems with quadratic stiff potentials, and they are *'preferable to symplectic methods for oscillatory differential equations'* [123]. On the other hand, we observed numerically that the long time performances of mollified impulse methods on the Fermi-Pasta-Ulam problem [101] (at the timescale $O(\omega)$, where ω is the fast frequency corresponding to the quadratic stiff potential) were less satisfactory than that of impulse methods [276]. In addition, neither mollified impulse methods nor Gautschi-type integrators can be viewed as splitting methods, and therefore it is not clear at this time how or whether it is possible to generalize them to integrate Langevin equations with stiff frictions using macroscopic timesteps (a way to generalize splitting methods to stiff Langevin equations is proposed in [276]).

IMEX is a variational integrator for stiff Hamiltonian systems [265]. It works by introducing a discrete Lagrangian via a trapezoidal approximation of the soft potential and a midpoint approximation of the fast potential. It is explicit in the case of quadratic stiff potential, but is implicit if the stiff potential has nonlinear derivatives. In addition to the drawback that implicit methods are usually slower than explicit methods if comparable step lengths are employed, there is no guarantee on IMEX' accuracy for general problems, because *'implicit methods in general fail to capture the effective dynamics of the slow time scale because they cannot correctly capture non-Dirac invariant distributions'* [184].

The second type of numerical algorithms, on the other hand, is based on a separation of slow or fast variables. Here is an incomplete list of examples:

The reversible averaging integrator proposed in [181] averages the force on slow variables and avoids resonant instabilities exhibited in impulse methods. It treats

the dynamics of slow and fast variables separately and assumes respectively piecewise linear and harmonically-oscillatory trajectories of the slow and fast variables. It is reversible, however, not symplectic.

In addition, a Hamilton-Jacobi approach is used to derive a homogenization method for multiscale Hamiltonian systems [176], and the resulting method is symplectic and works for not only quadratic but quasi-quadratic fast potentials. We also refer to [82] for a generalization of this method to systems that have either one slowly varying fast frequency or several constant frequencies. The difficulty with this analytical approach is how to deal with high-dimensional systems with different varying fast frequencies.

General multiscale mechanical systems: For general mechanical systems with non-quadratic stiff potentials, many different perspectives have been proposed.

Asynchronous Variational Integrators [183] use timesteps of different lengths to treat different extents of stiffness and provide a way to derive conservative symplectic integrators for PDEs. However, stiff potentials still require a fine time step discretization over the whole time evolution.

A similar idea is in the early work of multiple time-step methods [267], which evaluate forces to different extents of accuracies by approximating less important forces via Taylor expansions, but the idea has issues with long time behavior, stability and accuracy, as described in Section 5 of [180].

A popular method introduced by Fixman [103] is to freeze the fastest bond oscillations in polymers, so that stiffness in the equations could be removed. In order to correct the effect of freezing, a compensating log term analogous to an entropy-based free energy was added to the Hamiltonian. This method is successful in studying statistics of the system, but does not always reconstruct the correct dynamics [232, 227, 37].

Several homogenization approaches for Hamiltonian systems (in analogy to the classical homogenization theory [27, 158]) have been proposed. We refer to $\mathcal{M}$-

convergence introduced in [249, 38], the two-scale expansion of the solutions to the Hamilton-Jacobi form of Newton's equations with stiff quadratic potentials [176], and PDE methods in weak KAM theory [100]. We also refer to [59], [150], and [239].

In addition, it is worth mentioning that methods based on averaging, such as HMM and the equation-free method, could work for systems with arbitrary stiff potentials too; however, besides the difficulty in identifying the slow variable, as well as the necessity for using a smaller timestep (i.e., mesoscopic, as opposed to the macroscopic ones used by many of the above methods designed for quadratic stiff potentials), there has been no success in making these generic multiscale approaches symplectic for Hamiltonian systems. In their original form, these methods are based on the averaging of the instantaneous drifts of the slow variable, which breaks symplecticity in all variables. On the other hand, variants that preserve structures other than the symplecticity on all variables have been successfully proposed. By using Verlet/leap-frog macro-solvers, methods that are symplectic on slow variables (when those variables can be identified) have been proposed in the framework of HMM (Heterogeneous Multiscale Method) in [252, 56]. A 'reversible averaging' method has been proposed in [178] for mechanical systems with separated fast and slow variables. More recently, a reversible multiscale integration method for mechanical systems was proposed in [14] in the context of HMM. After tracking down the slow variables, this method enforces reversibility in all variables as an optimization constraint at each coarse step when minimizing the distance between the effective drift obtained from the micro-solver and the drift of the macro-solver. We also refer to [243] for a symmetric HMM for mechanical systems with stiff potentials of the form $\frac{1}{\epsilon}\sum_{j=1}^{\nu} g_j(q)^2$.

Why is symplectic integration good for multiscale Hamiltonian systems? Although backward error analysis (relating symplecticity and energy conservation) does not apply directly to stiff systems (due to large Lipschitz constants), improved long time behaviors of symplectic integrators, such as near-preservation of energy

and conservation of momentum maps, are often numerically observed. Modulated Fourier expansion [72] has been proposed to explain favorable long time energy behaviors of some integrators for oscillatory Hamiltonian systems.

Multiscale SDEs: Although a significant amount of research has been conducted in both the direction of geometric integration of SDEs (see, e.g., the discussion in Section 1.3; here 'geometric' means statistics capturing/preserving) and the direction of multiscale analysis/integration of SDEs (see, e.g., the discussion in Section 2.3.1), little has been done to combine the two, i.e., to study the geometric multiscale integration of SDEs.

Multiscale PDEs: Multiscale methods that provide conservative approximations have been proposed. For instance, [156] proposed a multiscale finite volume approach to elliptic problems, and the main idea is *'to use a finite volume global formulation with multiscale basis functions and obtain a mass conservative velocity field on a coarse grid'* [93]. *'A similar approach was independently proposed later'* [93] in [94], where a finite volume element method as *'a global coupling mechanism for multiscale basis functions'* [93] was formally introduced. The approach of [156] was generalized to parabolic problems [133]. Nevertheless, few conservative methods have been proposed for multiscale hyperbolic conservation laws.

To the best of my knowledge, so far there has been no multiscale multisymplectic integrators proposed for Hamiltonian PDEs.

1.4 Structure of this thesis

As can be seen from the above, the current status of the field (prior to this thesis) is as follows: (i) if a macroscopic integration step independent of ϵ is desired, existing symplectic multiscale integrators for mechanical systems are mostly limited to quadratic stiff potentials (with the exception of the Hamilton-Jacobi homogenization approach, which works for a subclass of quasi-quadratic stiff potentials), and they are not generalized (well enough) to stiff Langevin SDEs; (ii) to integrate

general multiscale ODEs, SDEs and PDEs using coarse steps, on the other hand, slow variables (depending on the situation, perhaps the fast ones as well) have to be identified; (iii) moreover, no multiscale ODE/PDE integrator that preserves symplecticity/multisymplecticity has been proposed. Additional open questions include: (iv) when a mechanical system with a quasi-quadratic stiff potential has a large amount of fast degrees of freedom, how can its high-dimensional frequency matrix be diagonalized in a symplectic and efficient way, if a diagonalization is necessary at all? (v) the simulation of constrained mechanical systems already uses a macroscopic timestep, but could it be made even faster?

(ii) and (iii) will be addressed in Chapters 2 and 3. Specifically, in Chapter 2, we propose a strategy to construct multiscale ODE/SDE integrators from arbitrary single-scale integrators. The resulting methods, called FLow AVeraging integratORs (FLAVORs), are two-scale flow convergent; more significantly, FLAVORs do not require any identification of slow or fast variables, and they inherit structure preservation properties from corresponding legacy codes, such as symplecticity.

In Chapter 3, the strategy of FLAVORization is extended to stiff PDEs. We show that various numerical PDE approaches, including finite difference, multisymplectic integrators, and pseudospectral methods, could all be FLAVORized (and hence made multiscale). Two-scale flow convergence of the numerical solutions can again be demonstrated.

Then, to address (i), we considered highly oscillatory mechanical systems in Chapter 4. The stiff potential is no longer limited to being quadratic (which corresponds to fast harmonic oscillators, and is generalized to stiff Langevin system and analyzed in Section 4.2), but instead allowed to be fully quasi-quadratic (i.e., fast 'harmonic' oscillators with a large number of distinct slowly varying frequencies, which are mixed by a slowly varying diagonalization frame; see Section 4.3). These treatments differ from the FLAVOR strategy for general multiscale systems, because the special context allows even faster computations (due to macroscopic integration steps) and better convergence properties (namely, strong convergence on both slow and fast positions).

An interesting by-product in Chapter 4 is the introduction of two numerical algebra algorithms: the first one (Section 4.3.4) is an efficient and generic method for the symplectic exponentiation of a matrix, and the second one (Section 4.3.5) is an efficient and generic method for repetitive (symplectic) exponentiations of a slowly varying sequence of matrices. These simple methods have highly nontrivial properties, and successfully address (iv).

In Chapter 5, we consider another special case of multiscale mechanical systems, in which the fast dynamics asymptotically approaches a Dirac point distribution as stiffness goes to infinity. For this case, it is again possible to employ a macroscopic integration step, for an implicit method is sufficient to capture the effective dynamics. The contribution is a method that replaces expensive nonlinear solves in the classical Newmark implicit integrators [214] by cheap linear solves, while yet the stability and symplecticity of these integrators are maintained.

An interesting application in this chapter is the cheap simulation of constrained dynamics, in which we replace rigid constraints by stiff springs oscillating around constrained values. Both speed and accuracy advantages over the classical constrained dynamics algorithm of SHAKE [237] are obtained. This provides an affirmative answer to (v).

In Chapter 6, we relieve the requirement on the accuracy of individual trajectories, and instead ask for an accuracy of the statistical properties of an ensemble of trajectories. This results in an accelerated approach for sampling an arbitrary statistical distribution, which is achieved by tuning the friction and annealing the temperature in the geometrical integration of a Langevin system. Besides the idea itself (surprisingly, little literature on the annealing idea applied to the problem of statistical sampling has been found), our contribution includes an analytical illustration of an optimal friction, a bound on the sampling error given a finite temperature cooling schedule, and a semi-empirical optimization of this bound. An interesting feature of this approach is that it could be used concurrently with many other accelerated sampling approaches, and the base Langvin integrator could be the multiscale ones mentioned above (although a rigorous theory on the

resulting geometric ergodicity has not been formulated yet), and in this way the speed-ups could be stacked.

Some other relevant topics are listed in Chapter 7. Section 7.1 concerns the integration of electric circuits, whose simulation is highly nontrivial because the system is constrained, degenerate, and subject to non-conservative forces. Two additional complications are: circuits are subject to environmental noise, and most modern circuits are multiscale. All these difficulties are solved in a variational framework, where constraints are handled by a projection, the degeneracy and forcing are dealt with in the framework of Lagrange-d'Alembert-Pontryagin principle and its discretization, the noise is treated by a stochastic variational principle, and the up-scaling is taken care of by an application of FLAVOR (Chapters 2 and 3). The physical implications of classical structure preserving properties of variational integrators are shown by co-authors (not included, see [218]), and a new preserved quantity of frequency spectrum is studied both numerically and analytically.

Section 7.2 describes a frequency domain approach for the efficient simulation of an acoustic wave in a nonlinear homogeneous medium (Westervelt equation). The ODE integrator in the frequency domain is essentially a first-order impulse method described in Section 4.2.1, which allows a macroscopic-time-stepping.

Section 7.3 proposes a quantification of the importance of mass effects in molecular dynamics. This is based on the optimization of the rate functional in Freidlin-Wentzell large deviation theory [107], which describes rates of transitions in SDEs. Two methods for the optimization are presented. The first is analytical, and it works with arbitrary starting and ending points, however only for linear systems. The second is numerical, and it works for arbitrary systems, however only with meta-stable starting and ending points.

Chapter 2

FLAVORs for ODEs and SDEs: Explicit geometric integrations of general stiff multiscale systems without the identification of slow or fast variables

To propose symplectic multiscale integrators for generic Hamiltonian systems without identifying the slow or fast variables, we designed FLow AVeraging integratORs (FLAVORs) [274]. FLAVORs are not restricted to Hamiltonian systems, but integrate general stiff multiscale ODEs and SDEs using a mesoscopic timestep, which means there is no need to resolve the fast timescale. Nevertheless, the correct slow dynamics will still be obtained. The idea is to account for the effective contribution of the fast variables by requiring the minimum amount of information on their dynamics, which turns out to be their local ergodic measure, and an average with respect to this measure can be approximated by averaging flow maps. This is very different from existing methods, such as HMM and the equation-free approach (see Section 1.3.2), all of which average instantaneous drifts.

Consequently, a FLAVOR can be constructed from any convergent single-scale legacy integrator. It inherits conservation properties (e.g., symplecticity) from the legacy method, and therefore provides the first symplectic approach to integrate multiscale mechanical systems. Moreover, FLAVORs do not require the fast or the slow timescale to be a priori identified, but only requires the existence of such a scale separation. In addition, a FLAVOR is explicit if the legacy method is

explicit. Unlike past methods reviewed in Section 1.3.2, FLAVORs apply in a unified way to both stiff systems with fast transients and stiff systems with rapid oscillations, with or without noise, using a mesoscopic integration timestep chosen independently from the stiffness. Because of all these, FLAVOR is the state of art method for accurate and efficient long time integrations of generic stiff multiscale systems.

Most of the results in this chapter are published in [274].

2.1 FLAVORs for general ODEs

2.1.1 Averaging

Consider the following ODE on $\mathbb{R}^d$,

$$\dot{u}^\epsilon = G(u^\epsilon) + \frac{1}{\epsilon} F(u^\epsilon). \tag{2.1}$$

In Subsections 2.1.8, 2.2.2, 2.3.2, 2.3.6 and 2.4.2 we will consider more general ODEs, stiff deterministic Hamiltonian systems (2.42), SDEs ((2.59) and (2.73)) and Langevin equations ((2.81) and (2.82)); however, for the sake of clarity, we will start the description of our method with (2.1).

Condition 2.1.1. *Assume that there exists a diffeomorphism $\eta := (\eta^x, \eta^y)$, from $\mathbb{R}^d$ onto $\mathbb{R}^{d-p} \times \mathbb{R}^p$ (with uniformly bounded C^1, C^2 derivatives), separating slow and fast variables, i.e., such that (for all $\epsilon > 0$) the process $(x_t^\epsilon, y_t^\epsilon) = (\eta^x(u_t^\epsilon), \eta^y(u_t^\epsilon))$ satisfies an ODE system of the form*

$$\begin{cases} \dot{x}^\epsilon = g(x^\epsilon, y^\epsilon) & x_0^\epsilon = x_0 \\ \dot{y}^\epsilon = \frac{1}{\epsilon} f(x^\epsilon, y^\epsilon) & y_0^\epsilon = y_0 \end{cases}. \tag{2.2}$$

Condition 2.1.2. *Assume that the fast variables in* (2.2) *are locally ergodic with respect to a family of measures μ drifted by slow variables. More precisely, we assume that there exists a family of probability measures $\mu(x, dy)$ on $\mathbb{R}^p$ indexed*

by $x \in \mathbb{R}^{d-p}$ and a positive function $T \mapsto E(T)$ such that $\lim_{T\to\infty} E(T) = 0$ and such that for all x_0, y_0, T and ϕ uniformly bounded and Lipschitz, the solution to

$$\dot{Y}_t = f(x_0, Y_t) \qquad Y_0 = y_0 \tag{2.3}$$

satisfies

$$\Big|\frac{1}{T}\int_0^T \phi(Y_s)ds - \int_{\mathbb{R}^p} \phi(y)\mu(x_0, dy)\Big| \leq \chi\big(\|(x_0, y_0)\|\big)E(T)(\|\phi\|_{L^\infty} + \|\nabla\phi\|_{L^\infty}) \tag{2.4}$$

where $r \mapsto \chi(r)$ is bounded on compact sets.

Under Conditions 2.1.1 and 2.1.2, it is known (we refer for instance to [239] or to Theorem 14, Section 3 of Chapter II of [258] or to [226]) that x^ϵ converges towards x_t defined as the solution to the ODE

$$\dot{x} = \int g(x, y)\mu(x, dy), \quad x|_{t=0} = x_0 \tag{2.5}$$

where $\mu(x, dy)$ is the ergodic measure associated with the solution to the ODE

$$\dot{y} = f(x, y), \tag{2.6}$$

in which the slow variable x is fixed.

It follows that the slow behavior of solutions of (2.1) can be simulated over coarse time steps by first identifying the slow process x^ϵ and then using numerical approximations of solutions of (2.2) to approximate x^ϵ. At least two classes of integrators have been founded on this observation: The equation free method [165, 164] and the Heterogeneous Multiscale Method [86, 97, 85, 13]. One shared characteristic of the original form of those integrators is, after identification of the slow variables, to use a micro-solver to approximate the effective drift in (2.5) by averaging the instantaneous drift g with respect to numerical solutions of (2.6) over a time span larger than the mixing time of the solution to (2.6).

2.1.2 FLAVORs

Instead of averaging the instantaneous drift on the slow variable x^ϵ (the first equation in (2.2)) with respect to samples of the fast variable y^ϵ, we propose to average the instantaneous flow of the ODE (2.1) with the slow and fast variables hidden. We call the resulting class of numerical integrators FLow AVeraging integratORS (FLAVORS). Since FLAVORS are directly applied to (2.1), hidden slow variables do not need to be identified, either explicitly or numerically. Furthermore FLAVORS can be implemented using an arbitrary legacy integrator $\Phi_h^{\frac{1}{\epsilon}}$ for (2.1) in which the parameter $\frac{1}{\epsilon}$ can be controlled (Figure 2.1). More precisely, assume that

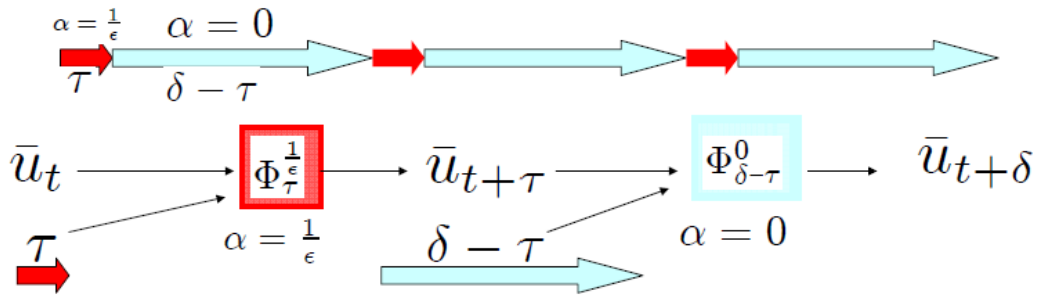


Figure 2.1: A pre-existing numerical scheme resolving the microscopic time scale can be used as a black box and turned into a FLAVOR by simply turning on and off stiff parameters over a microscopic timescale τ (on) and a mesoscopic timescale δ (off). The bottom line of the approach is to (repeatedly) compose an accurate, short-time integration of the complete set of equations with an accurate, intermediate-time integration of the non-stiff part of the system. While the integration over short time intervals is accurate (in a strong sense), this is extended to intermediate time integration (in the sense of measures) using the interplay between the short time integration and the mesoscopic integration. The computational cost remains bounded independently from the stiff parameter $1/\epsilon$ because: (i) The whole system is only integrated over an extremely short ($\tau \ll \epsilon$) time interval during every intermediate (δ) time interval. (ii) The intermediate time step δ (that of the non-stiff part of the system) is limited not by the fast time scales (ϵ) but by the slow ones ($\mathcal{O}(1)$).

there exists a constant $h_0 > 0$ such that Φ_h^α satisfies for all $h \leq h_0 \min(\frac{1}{\alpha}, 1)$ and $u \in \mathbb{R}^d$

$$\left|\Phi_h^\alpha(u) - u - hG(u) - \alpha hF(u)\right| \leq Ch^2(1+\alpha)^2 \tag{2.7}$$

then FLAVOR can be defined as the algorithm simulating the process

$$\bar{u}_t = \left(\Phi^0_{\delta-\tau} \circ \Phi^{\frac{1}{\epsilon}}_{\tau}\right)^k (u_0) \quad \text{for} \quad k\delta \leq t < (k+1)\delta \tag{2.8}$$

where τ is a fine time step resolving the fast time scale ($\tau \ll \epsilon$) and δ is a mesoscopic time step independent of the fast time scale satisfying $\tau \ll \epsilon \ll \delta \ll 1$ and

$$(\frac{\tau}{\epsilon})^2 \ll \delta \ll \frac{\tau}{\epsilon} \tag{2.9}$$

In our numerical experiments, we have used the "rule of thumb" $\delta \sim \gamma \frac{\tau}{\epsilon}$ where γ is a small parameter (0.1 for instance).

By switching stiff parameters FLAVOR approximates the flow of (2.1) over a coarse time step h (resolving the slow time scale) by the flow

$$\Phi_h := \left(\Phi^0_{\frac{h}{M}-\tau} \circ \Phi^{\frac{1}{\epsilon}}_{\tau}\right)^M \tag{2.10}$$

where M is a positive integer corresponding to the number of "samples" used to average the flow (δ has to be identified with $\frac{h}{M}$). We refer to Section 2.1.5 for the distinction between macro- and meso-steps, the intuition behind the timesteps requirement (2.9), and the rationale and mechanism behind FLAVORs.

Since FLAVORs are obtained by flow-composition, we will show in Section 2.2 and 2.4 that they inherit the structure preserving properties (for instance symplecticity and symmetries under a group action) of the legacy integrator for Hamiltonian systems and Langevin equations.

Under conditions (2.9) on τ and δ, we show that (2.8) is strongly accurate with respect to (hidden) slow variables and weakly (in the sense of measures) accurate with respect to (hidden) fast variables. Motivated by this observation, we introduce the related notion of **two-scale flow convergence** in analogy with homogenization theory for elliptic PDEs [215, 4] and call it F-convergence for short. F-convergence is close in spirit to the Young measure approach to computing slowly advancing fast oscillations introduced in [18, 17].

2.1.3 Two-scale flow convergence

Let $(\xi_t^\epsilon)_{t\in\mathbb{R}^+}$ be a sequence of processes on $\mathbb{R}^d$ (functions from $\mathbb{R}^+$ to $\mathbb{R}^d$) indexed by $\epsilon > 0$. Let $(X_t)_{t\in\mathbb{R}^+}$ be a process on $\mathbb{R}^{d-p}$ ($p \geq 0$). Let $x \mapsto \nu(x, dz)$ be a function from $\mathbb{R}^{d-p}$ into the space of probability measures on $\mathbb{R}^d$.

Definition 2.1.1. *We say that the process ξ_t^ϵ F-converges to $\nu(X_t, dz)$ as $\epsilon \downarrow 0$ and write $\xi_t^\epsilon \xrightarrow[\epsilon\to 0]{F} \nu(X_t, dz)$ if and only if for all functions φ bounded and uniformly Lipschitz-continuous on $\mathbb{R}^d$, and for all $t > 0$,*

$$\lim_{h\to 0}\lim_{\epsilon\to 0}\frac{1}{h}\int_t^{t+h}\varphi(\xi_s^\epsilon)\,ds = \int_{\mathbb{R}^d}\varphi(z)\nu(X_t, dz) \tag{2.11}$$

The idea is that X is the slow variable, and $\nu(X_t, dz)$ corresponds to a measure on the full space (including both the slow and the fast variables) for a given X_t. For the case of FLAVORs, $\nu(X_t, dz)$ will correspond to a Dirac distribution concentrated at the value of the slow variable X_t, times the local ergodic measure of the fast variable, and then pulled back to the original coordinates by the scale separation diffeomorphism.

2.1.4 Asymptotic convergence result

Our convergence theorem requires that u_t^ϵ and $\bar{u}_t$ do not blow up as $\epsilon \downarrow 0$; more precisely, we will assume that the following conditions are satisfied:

Condition 2.1.3.

1. *F and G are Lipschitz continuous.*
2. *For all u_0, $T > 0$, the trajectories $(u_t^\epsilon)_{0\leq t\leq T}$ are uniformly bounded in ϵ.*
3. *For all u_0, $T > 0$, the trajectories $(\bar{u}_t^\epsilon)_{0\leq t\leq T}$ are uniformly bounded in ϵ, $0 < \delta \leq h_0$, $\tau \leq \min(h_0\epsilon, \delta)$.*

For π, an arbitrary measure on $\mathbb{R}^d$, we define $\eta^{-1} * \pi$ to be the push forward of the measure π by η^{-1}.

Theorem 2.1.1. *Let u_t^ϵ be the solution to* (2.1) *and $\bar{u}_t$ be defined by* (2.8)*. Assume that equation* (2.7) *and Conditions 2.1.1, 2.1.2 and 2.1.3 are satisfied, then*

- u_t^ϵ *F-converges to* $\eta^{-1} * \big(\delta_{X_t} \otimes \mu(X_t, dy)\big)$ *as* $\epsilon \downarrow 0$ *where* X_t *is the solution to*

$$\dot{X}_t = \int g(X_t, y)\, \mu(X_t, dy) \qquad X_0 = x_0. \tag{2.12}$$

- $\bar{u}_t$ *F-converges to* $\eta^{-1} * \big(\delta_{X_t} \otimes \mu(X_t, dy)\big)$ *for* $\epsilon \leq \delta/(-C \ln \delta)$, $\frac{\tau}{\epsilon} \downarrow 0$, $\frac{\epsilon}{\tau}\delta \downarrow 0$ *and* $(\frac{\tau}{\epsilon})^2 \frac{1}{\delta} \downarrow 0$.

We refer to Section A.1 of the appendix for the detailed proof of Theorem 2.1.1.

Remark 2.1.1. *The F-convergence of* u_t^ϵ *to* $\eta^{-1} * \big(\delta_{X_t} \otimes \mu(X_t, dy)\big)$ *can be restated as*

$$\lim_{h \to 0} \lim_{\epsilon \to 0} \frac{1}{h} \int_t^{t+h} \varphi(u_s^\epsilon)\, ds = \int_{\mathbb{R}^p} \varphi(\eta^{-1}(X_t, y)) \mu(X_t, dy) \tag{2.13}$$

for all functions φ *bounded and uniformly Lipschitz-continuous on* $\mathbb{R}^d$*, and for all* $t > 0$.

Remark 2.1.2. *Observe that* g *comes from* (2.5). *It is not explicitly known and does not need to be explicitly known for the implementation of the proposed method.*

Remark 2.1.3. *The limits on* ϵ, τ *and* δ *are in essence stating that FLAVOR is accurate provided that* $\tau \ll \epsilon$ *(*τ *resolves the stiffness of* (2.1)*) and equation* (2.9) *is satisfied.*

Remark 2.1.4. *Throughout this chapter,* C *will refer to an appropriately large enough constant independent from* ϵ, δ, τ. *To simplify the presentation of our results, we use the same letter* C *for expressions such as* $2Ce^C$ *instead of writing it as a new constant* C_1 *independent from* ϵ, δ, τ.

2.1.5 Rationale and mechanism behind FLAVORs

We will now explain the rationale and mechanism behind FLAVORs. Let us start by considering the case where η is the identity diffeomorphism. Let $\varphi^{\frac{1}{\epsilon}}$ be the flow of (2.2). Observe that φ^0 (obtained from $\varphi^{\frac{1}{\epsilon}}$ by setting the parameter $\frac{1}{\epsilon}$ to zero) is the flow of (2.2) with y^ϵ frozen, i.e.,

$$\varphi^0(x, y) = (\hat{x}_t, y) \quad \text{where } \hat{x}_t \text{ solves } \quad \frac{d\hat{x}}{dt} = g(\hat{x}, y), \quad \hat{x}_0 = x. \tag{2.14}$$

The main effect of FLAVORs is to average the flow of (2.2) with respect to fast degrees of freedom via splitting and re-synchronization. By splitting, we refer to the substitution of the flow $\varphi_\delta^{\frac{1}{\epsilon}}$ by composition of $\varphi_{\delta-\tau}^0$ and $\varphi_\tau^{\frac{1}{\epsilon}}$, and by re-synchronization we refer to the distinct time-steps δ and τ whose effects are to advance the internal clock of fast variables by τ every step of length δ. By averaging, we refer to the fact that FLAVORs approximates the flow $\varphi_H^{\frac{1}{\epsilon}}$ by the flow

$$\varphi_H := \left(\varphi^0_{\frac{H}{M}-\tau} \circ \varphi_\tau^{\frac{1}{\epsilon}}\right)^M \tag{2.15}$$

where H is a macroscopic time step resolving the slow timescale associated with x^ϵ, M is a positive integer corresponding to the number of samples used to average the flow (δ is identified with $\frac{H}{M}$), and τ is a microscopic time step resolving the fast timescale, of the order of ϵ, and associated with y^ϵ. In general, analytical formulae are not available for φ^0 and $\varphi^{\frac{1}{\epsilon}}$, and numerical approximations are used instead.

Observe that when FLAVORs are applied to systems with explicitly separated slow and fast processes, they lead to integrators that are locally in the neighborhood of those obtained with HMM (or equation-free) with a reinitialization of the fast variables at macrotime n by their final value at macrotime step $n-1$ and with only one microstep per macrostep [87, 89].

We will now consider the situation where η is not the identity map and give the rationale behind the step size requirements (2.9).

$$\begin{array}{ccccc}
\bar{u}_{n\delta} & \xrightarrow{\Phi_\tau^{\frac{1}{\epsilon}}} & \bar{u}_{n\delta+\tau} & \xrightarrow{\Phi^0_{\delta-\tau}} & \bar{u}_{(n+1)\delta} \\
{\scriptstyle\eta^{-1}}\uparrow\downarrow{\scriptstyle\eta} & & {\scriptstyle\eta^{-1}}\uparrow\downarrow{\scriptstyle\eta} & & {\scriptstyle\eta^{-1}}\uparrow\downarrow{\scriptstyle\eta} \\
(\bar{x},\bar{y})_{n\delta} & \xrightarrow{\Psi_\tau^{\frac{1}{\epsilon}}} & (\bar{x},\bar{y})_{n\delta+\tau} & \xrightarrow{\Psi^0_{\delta-\tau}} & (\bar{x},\bar{y})_{(n+1)\delta}
\end{array}$$

As illustrated in the above diagram, since $(\bar{x}_t, \bar{y}_t) = \eta(\bar{u}_t)$, simulating $\bar{u}_{n\delta}$ defined in (2.8) is equivalent to simulating the discrete process

$$(\bar{x}_{n\delta}, \bar{y}_{n\delta}) := \left(\Psi^0_{\delta-\tau} \circ \Psi_\tau^{\frac{1}{\epsilon}}\right)^n (x_0, y_0) \tag{2.16}$$

where

$$\Psi_h^\alpha := \eta \circ \Phi_h^\alpha \circ \eta^{-1} \tag{2.17}$$

Observe that the accuracy (in the topology induced by F-convergence) of $\bar{u}_t$ with respect to u_t^ϵ, solution of (2.1), is equivalent to that of $(\bar{x}_t, \bar{y}_t)$ with respect to $(x_t^\epsilon, y_t^\epsilon)$ defined by (2.2). Now, for the clarity of the presentation, assume that

$$\Phi_h^\alpha(u) = u + hG(u) + \alpha hF(u) \tag{2.18}$$

Using Taylor's theorem and (2.18), we obtain that

$$\Psi_h^\alpha(x,y) = (x,y) + h\big(g(x,y),0\big) + \alpha h\big(0, f(x,y)\big) + \int_0^1 v^T \operatorname{Hess}\eta(u+tv)v(1-t)^2\,dt \tag{2.19}$$

with

$$u := \eta^{-1}(x,y) \quad \text{and} \quad v := h(G+\alpha F)\circ\eta^{-1}(x,y) \tag{2.20}$$

It follows from (2.19) and (2.20) that $\Psi_h^{\frac{1}{\epsilon}}$ is a first-order-accurate integrator approximating the flow of (2.2) and Ψ_h^0 is a first-order-accurate integrator approximating the flow of (2.14). Let H be a coarse time step and δ a meso-step. Since $\bar{x}$ remains nearly constant over the coarse time step, the switching (on and off) of the stiffness parameter $\frac{1}{\epsilon}$ averages the drift g of $\bar{x}$ with respect to the trajectory of $\bar{y}$ over H. Since the coarse step H is composed of $\frac{H}{\delta}$ mesosteps, the internal clock of the fast process is advanced by $\frac{H}{\delta} \times \frac{\tau}{\epsilon}$. Since $H = \mathcal{O}(1)$, the trajectory of $\bar{y}$ is mixing with respect to the local ergodic measure μ provided that $\frac{\tau}{\delta\epsilon} \gg 1$, i.e.,

$$\delta \ll \frac{\tau}{\epsilon} \tag{2.21}$$

Equation (2.21) corresponds to the right hand side of equation (2.9). If η is a non-linear diffeomorphism (with non-zero Hessian), it also follows from equations (2.19) and (2.20) that each invocation of the integrator $\Psi_\tau^{\frac{1}{\epsilon}}$ occasions an error (on the accuracy of the slow process) proportional to $(\frac{\tau}{\epsilon})^2$. Since during the coarse time step H, $\Psi_\tau^{\frac{1}{\epsilon}}$ is called $\frac{H}{\delta}$-times, it follows that the error accumulation during

H is $\frac{H}{\delta} \times (\frac{\tau}{\epsilon})^2$. Hence, the accuracy of the integrator requires that $\frac{1}{\delta} \times (\frac{\tau}{\epsilon})^2 \ll 1$, i.e.,

$$\left(\frac{\tau}{\epsilon}\right)^2 \ll \delta \tag{2.22}$$

Equation (2.22) corresponds to the left hand side of equation (2.9).

Observe that if η is linear, its Hessian is null and the remainder on the right hand side of (2.19) is zero. It follows that if η is linear, the error accumulation due to fine time steps on slow variables is zero and Condition (2.21) is sufficient for the accuracy of the integrator.

It has been observed in [88] and in Section 5 of [289] that slow variables do not need to be identified with HMM/averaging type integrators if η is a linear map and

$$\frac{\Delta t}{M} \ll \frac{\tau}{\epsilon} \tag{2.23}$$

where is M the number of fine-step iterations used by HMM to compute the average the drift of slow variables and Δt is the coarse time step (in HMM) along the direction of the averaged drift. The analysis of FLAVORs associated with equation (2.19) reaches a similar conclusion that if η is linear in the sense that the error caused by the Hessian of η in (2.19) is zero then the (sufficient) condition (2.21) is analogous to (2.23) for $M = 1$. It is also stated on Page 2 of [88] that *"there are counterexamples showing that algorithms of the same spirit do not work for deterministic ODEs with separated time scales if the slow variables are not explicitly identified and made use of. But in the present context, the slow variables are linear functions of the original variables, and this is the reason why the seamless algorithm works."* Here, the analysis of FLAVORs associated with equation (2.19) shows an algorithm based on an averaging principle would indeed, in general, not work if η is nonlinear (and (2.22) not satisfied) due to the error accumulation (on slow variables) associated with the Hessian of η. However, the above analysis also shows that if Condition (2.22) is satisfied, then, although η may be nonlinear, FLAVORs will always work without the identification of the slow variables.

2.1.6 Non-asymptotic convergence result

Theorem 2.1.2. *Under assumptions and notations of Theorem 2.1.1, there exists $C > 0$ such that for $\delta < h_0$, $\tau < h_0\epsilon$ and $t > 0$,*

$$|x_t^\epsilon - \eta^x(\bar{u}_t)| \leq Ce^{Ct}\chi_1(u_0, \epsilon, \delta, \tau) \tag{2.24}$$

and

$$\left|\frac{1}{T}\int_t^{t+T} \varphi(\bar{u}_s)\, ds - \int_{\mathbb{R}^p} \varphi(\eta^{-1}(X_t, y))\mu(X_t, dy)\right| \leq \chi_2(u_0, \epsilon, \delta, \tau, T, t)(\|\varphi\|_{L^\infty} + \|\nabla\varphi\|_{L^\infty}) \tag{2.25}$$

where χ_1 and χ_2 are functions converging towards zero as $\epsilon \leq \delta/(C\ln\frac{1}{\delta})$, $\frac{\tau}{\epsilon} \downarrow 0$, $\frac{\epsilon}{\tau}\delta \downarrow 0$ and $(\frac{\tau}{\epsilon})^2\frac{1}{\delta} \downarrow 0$ (and $T \downarrow 0$ for χ_2).

Remark 2.1.5. *For $\epsilon \leq \delta/(-C\ln\delta)$ and $\delta\frac{\epsilon}{\tau} + \frac{\tau}{\epsilon} \leq 1$, the following holds*

$$\chi_1(u_0, \epsilon, \delta, \tau) \leq \sqrt{\delta} + \big(\frac{\tau}{\epsilon}\big)^2\frac{1}{\delta} + E\big(\frac{1}{C}\ln\frac{1}{\delta}\big) + \big(\frac{\delta\epsilon}{\tau}\big)^{\frac{1}{2}} + \big(\frac{\tau}{\epsilon}\big)^{\frac{1}{2}} + E\Big(\frac{1}{C}\ln\Big(\big(\frac{\delta\epsilon}{\tau} + \frac{\tau}{\epsilon}\big)^{-1}\Big)\Big) \tag{2.26}$$

and χ_2 satisfies a similar inequality.

Remark 2.1.6. *Choosing $\tau \sim \gamma\epsilon$ and $\delta \sim \gamma\frac{\tau}{\epsilon}$,where γ is a small constant independent from ϵ, Theorem 2.1.2 shows that the approximation error of FLAVOR is bounded by a function of γ converging towards zero as $\gamma \downarrow 0$. If follows that the speed up is of the order of $\frac{\delta}{\tau} \sim \frac{\gamma}{\epsilon}$, i.e., scales like $\frac{1}{\epsilon}$ at fixed accuracy. In order to be able to compare FLAVOR with integrators resolving the fast timescale using fine time steps, we have limited ϵ from being too small, and hence the speed up in the numerical experiments to $200\times$ (but this can be arbitrary large as $\epsilon \downarrow 0$). For sufficiently small ϵ, we observe that FLAVORs with microstep τ and mesostep δ overperform their associated legacy integrator with the same microstep τ over large simulation times (we refer to Section 2.6.3 on the Fermi-Pasta-Ulam problem). This phenomenon is caused by an error accumulation at each tick (microstep) of*

the clock of fast variables. Since FLAVORs (indirectly, i.e., without identifying fast variables) slow down the speed of this clock from $\frac{1}{\epsilon}$ *to a value* $\frac{\tau}{\delta\epsilon} \sim \frac{1}{\gamma}$ *independent from* ϵ*, this error does not blow up as* $\epsilon \downarrow 0$ *(as opposed to for an integrator that resolves the fast timescale). For this reason, if this error accumulation on fast variables is exponential, then the speed up at fixed accuracy does not scale like* $\frac{1}{\epsilon}$*, but like* $e^{\frac{T}{\epsilon}}$ *where* T *is the total simulation time. A consequence of this phenomenon can be seen in Figure 2.10 (associated with the FPU problem) where Velocity-Verlet fails to capture the* $\mathcal{O}(\epsilon^{-1})$ *dynamics with a time step* $h = 10^{-5}$ *whereas FLAVORs remain accurate with* $\tau = 10^{-4}$ *and* $\delta = 2 \cdot 10^{-3}$*.*

Remark 2.1.7. *The reader should not be surprised by the presence of the exponential factor* e^{Ct} *in* (2.24)*. It is known that global errors for numerical approximations of ODEs grow, in general, exponentially with time (see for instance [129]). These bounds are, however, already tight; consider, for instance, how error propagates in a generic numerical scheme applied to the special system of* $\dot{x} = x$*. It is possible to show that the increase of global errors is linear in time only for a restricted class of ODEs (using techniques from Lyapunov's theory of stability [293]). Notice that the constant* C *in the exponential of our bound does not scale with* ϵ^{-1}*, and therefore the bound is uniform and rather tight.*

Remark 2.1.8. *We refer to [97] for higher order averaging based methods. In particular, [97] shows how, after identification of slow variables, balancing the different error contributions yields an explicit stable integration method having the order of the macro scheme.*

2.1.7 Natural FLAVORs

Although convenient, it is not necessary to use legacy integrators to obtain FLAVORs. More precisely, Theorems 2.1.1 and 2.1.2 remain valid if FLAVORs are defined to be algorithms simulating the discrete process

$$\bar{u}_t := \left(\theta^G_{\delta-\tau} \circ \theta^\epsilon_\tau\right)^k (u_0) \quad \text{for} \quad k\delta \leq t < (k+1)\delta \tag{2.27}$$

where θ^ϵ_τ and $\theta^G_{\delta-\tau}$ are two mappings from $\mathbb{R}^d$ onto $\mathbb{R}^d$ (the former approximating the flow of the whole system (2.1) for time τ, and the latter approximating the flow of $\dot{v} = G(v)$ for time $\delta - \tau$), satisfying the following conditions:

Condition 2.1.4. 1. *There exists $h_0, C > 0$ such that for $h \leq h_0$ and any $u \in \mathbb{R}^d$,*

$$\left|\theta^G_h(u) - u - hG(u)\right| \leq Ch^2 \tag{2.28}$$

2. *There exists $\tau_0, C > 0$, such that for $\frac{\tau}{\epsilon} \leq \tau_0$ and any $u \in \mathbb{R}^d$,*

$$\left|\theta^\epsilon_\tau(u) - u - \tau G(u) - \frac{\tau}{\epsilon}F(u)\right| \leq C\left(\frac{\tau}{\epsilon}\right)^2 \tag{2.29}$$

3. *For all u_0, $T > 0$, the discrete trajectories $\left(\left(\theta^G_{\delta-\tau} \circ \theta^\epsilon_\tau\right)^k(u_0)\right)_{0\leq k\leq T/\delta}$ are uniformly bounded in ϵ, $0 < \delta \leq h_0$, $\tau \leq \min(\tau_0\epsilon, \delta)$.*

Observe that (2.8) is a particular case of (2.27) in which $\theta^\epsilon = \Phi^{\frac{1}{\epsilon}}$ and the mapping θ^G is obtained from the legacy integrator Φ^α by setting α to zero. We sometimes call (2.8) a nonintrusive FLAVOR for distinction.

2.1.8 FLAVORs for generic stiff ODEs

FLAVORs have a natural generalization to systems of the form

$$\dot{u}^{\alpha,\epsilon} = F(u^{\alpha,\epsilon}, \alpha, \epsilon) \tag{2.30}$$

where $u \mapsto F(u, \alpha, \epsilon)$ is Lipschitz continuous.

Condition 2.1.5. *Assume that:*

1. *$\epsilon \mapsto F(u, \alpha, \epsilon)$ is uniformly continuous in the neighborhood of 0.*

2. *There exists a diffeomorphism $\eta := (\eta^x, \eta^y)$, from $\mathbb{R}^d$ onto $\mathbb{R}^{d-p} \times \mathbb{R}^p$, independent from ϵ, α, with uniformly bounded C^1, C^2 derivatives, such that the process $(x^\alpha_t, y^\alpha_t) = \left(\eta^x(u^{\alpha,0}_t), \eta^y(u^{\alpha,0}_t)\right)$ satisfies, for all $\alpha \geq 1$, the ODE*

$$\dot{x}^\alpha = g(x^\alpha, y^\alpha) \quad x^\alpha_0 = x_0, \tag{2.31}$$

where $g(x,y)$ is Lipschitz continuous in x and y on bounded sets.

3. *There exists $\nu > 0$ and a family of probability measures $\mu(x,dy)$ on $\mathbb{R}^p$ such that for all x_0, y_0, T $\big((x_0,y_0) := \eta(u_0)\big)$ and φ uniformly bounded and Lipschitz*

$$\Big|\frac{1}{T}\int_0^T \varphi(y_s^\alpha)\,ds - \int_{\mathbb{R}^p}\varphi(y)\mu(x_0,dy)\Big| \le \chi\big(\|(x_0,y_0)\|\big)\big(E_1(T)+E_2(T\alpha^\nu)\big)\|\nabla\varphi\|_{L^\infty} \tag{2.32}$$

where $r \mapsto \chi(r)$ is bounded on compact sets and $E_2(r) \to 0$ as $r \to \infty$ and $E_1(r) \to 0$ as $r \to 0$.

4. *For all u_0, $T > 0$, the trajectories $(u_t^{\alpha,0})_{0\le t\le T}$ are uniformly bounded in $\alpha \ge 1$.*

Remark 2.1.9. *Observe that slow variables are not kept frozen in equation* (2.32)*. The error on local invariant measures induced by the (slow) drift of x^α is controlled by E_2. More precisely, the convergence of the right hand side of* (2.32) *towards zero requires that T goes to zero and at the same time $T\alpha^\nu$ goes towards infinity.*

Assume that we are given a mapping $\Phi_h^{\alpha,\epsilon}$ from $\mathbb{R}^d$ onto $\mathbb{R}^d$ approximating the flow of (2.30). If the parameter α can be controlled, then $\Phi_h^{\alpha,\epsilon}$ can be used as a black box for accelerating the computation of solutions of (2.30).

Condition 2.1.6. *Assume that:*

1. *There exists a constant $h_0 > 0$ such that $\Phi^{\alpha,\epsilon}$ satisfies for all $h \le h_0 \min(\frac{1}{\alpha^\nu}, 1)$, $0 < \epsilon \le 1 \le \alpha$*

$$\big|\Phi_h^{\alpha,\epsilon}(u) - u - hF(u,\alpha,\epsilon)\big| \le C(u)h^2(1+\alpha^{2\nu}) \tag{2.33}$$

where $C(u)$ is bounded on compact sets.

2. *For all u_0, $T > 0$, the discrete trajectories $\Big(\big(\Phi_{\delta-\tau}^{0,\epsilon}\circ\Phi_\tau^{\frac{1}{\epsilon},\epsilon}\big)^k(u_0)\Big)_{0\le k\le T/\delta}$ are uniformly bounded in $0 < \epsilon \le 1$, $0 < \delta \le h_0$, $\tau \le \min(h_0\epsilon^\nu, \delta)$.*

FLAVOR can be defined as the algorithm given by the process

$$\bar{u}_t = \left(\Phi_{\delta-\tau}^{0,\epsilon} \circ \Phi_{\tau}^{\frac{1}{\epsilon},\epsilon}\right)^k (u_0) \quad \text{for} \quad k\delta \leq t < (k+1)\delta \tag{2.34}$$

The theorem below shows the accuracy of FLAVORs for $\delta \ll h_0$, $\tau \ll \epsilon^\nu$ and $\left(\frac{\tau}{\epsilon^\nu}\right)^2 \ll \delta \ll \frac{\tau}{\epsilon^\nu}$.

Theorem 2.1.3. *Let $u_t^{\frac{1}{\epsilon},\epsilon}$ be the solution to (2.30) with $\alpha = 1/\epsilon$ and $\bar{u}_t$ be defined by (2.34). Assume that Conditions 2.1.5 and 2.1.6 are satisfied then*

- *$u_t^{\frac{1}{\epsilon},\epsilon}$ F-converges towards $\eta^{-1} * \left(\delta_{X_t} \otimes \mu(X_t, dy)\right)$ as $\epsilon \downarrow 0$ where X_t is the solution to*

$$\dot{X}_t = \int_{\mathbb{R}^p} g(X_t, y)\, \mu(X_t, dy) \qquad X_0 = x_0, \tag{2.35}$$

- *As $\epsilon \downarrow 0$, $\tau\epsilon^{-\nu} \downarrow 0$, $\delta\frac{\epsilon^\nu}{\tau} \downarrow 0$, $\frac{\tau^2}{\epsilon^{2\nu}\delta} \downarrow 0$, $\bar{u}_t$ F-converges towards $\eta^{-1} * \left(\delta_{X_t} \otimes \mu(X_t, dy)\right)$ as $\epsilon \downarrow 0$ where X_t is the solution of (2.35).*

Proof. The proof of Theorem 2.1.3 is similar to that of Theorem 2.1.1 and 2.3.1. Only the idea of the proof will be given here. The condition $\epsilon \ll 1$ is needed for the approximation of $u^{\alpha,\epsilon}$ by $u^{\alpha,0}$ and for the F-convergence of $u^{\frac{1}{\epsilon},0}$. Since $y_t^\alpha = \eta^y(u_t^{\alpha,0})$ the condition $\tau \ll \epsilon^\nu$ is used along with equation (2.33) for the accuracy of $\Phi_\tau^{\frac{1}{\epsilon},\epsilon}$ in (locally) approximating y_t^α. The condition $\delta \ll \frac{\tau}{\epsilon^\nu}$ allows for the averaging of g to take place prior to a significant change of x_t^α; more precisely, it allows for $m \gg 1$ iterations of $\Phi_\tau^{\frac{1}{\epsilon},\epsilon}$ prior to a significant change of x_t^α. The condition $\left(\frac{\tau}{\epsilon^\nu}\right)^2 \ll \delta$ is required in order to control the error accumulated by m iterations of $\Phi_\tau^{\frac{1}{\epsilon},\epsilon}$. □

Remark 2.1.10. *It is easy to see that Theorem 2.1.3 remains valid if Item 4 of Condition 2.1.5 and Item 2 of Condition 2.1.6 do not hold for all u_0 but only for a subset of initial conditions $u_0 \in \mathcal{I}$ for some $\mathcal{I} \subset \mathbb{R}^d$ and the trajectories of u and $\bar{u}$ remain in $\mathcal{I}$ for all ϵ.*

We also observe that Theorem 2.1.3 can easily be generalized to situations where η is not injective, for instance to a situation where η is a differentiable mapping

from $\mathbb{R}^d$ onto $\mathbb{R}^{d-p} \times \mathbb{R}^q$ where $q < p$. In that situation, Item 4 of Condition 2.1.5 and Item 2 of Conditional 2.1.6 should be replaced by the condition that $\eta(u)$ and $\eta(\bar{u})$ do not blow up as $\epsilon \downarrow 0$. Furthermore, the convergence of u and $\bar{u}$ are only partial in the sense that $\eta(u)$ F-converges towards $\delta_{X_t} \otimes \mu(X_t, dy)$ but the projection of u on the kernel of η (i.e., $\eta^{-1}(0,0)$) may not F-converge.

2.1.9 Limitations of the method

The proof of the accuracy of the method (Theorems 2.1.1 and 2.1.2) is based on an averaging principle; hence, if ϵ is not small (the stiffness of the ODE is weak), although the method may be stable, there is no guarantee of accuracy. More precisely, the global error of the method is an increasing function of ϵ, δ, $\frac{\tau}{\epsilon}$, $\frac{\delta\epsilon}{\tau}$, $(\frac{\tau}{\epsilon})^2\delta$. Writing $\gamma := \frac{\tau}{\epsilon}$, the accuracy requires $\gamma^2 \ll \delta \ll \gamma$. Choosing $\delta = \gamma^{\frac{3}{2}}$, the condition $\epsilon \ll \delta \ll 1$ (related to computational gain) requires $\epsilon^{\frac{2}{3}} \ll \gamma \ll 1$, which can be satisfied only if ϵ is small.

The other limitation of the method lies in the fact that a stiff parameter $\frac{1}{\epsilon}$ needs to be clearly identified. In many examples of interest (Navier-Stokes equations, Maxwell's equations,...), stiffness is a result of nonlinearity, initial conditions or boundary conditions and not of the existence of a large parameter $\frac{1}{\epsilon}$. Molecular dynamics can also create widely separated time-scales from nonlinear effects; we refer, for instance, to [301] and references therein.

2.2 FLAVORs for deterministic mechanical systems

2.2.1 Hamiltonian system and its geometric integration

Deterministic mechanical systems are governed by Hamiltonian equations [3]. Since averaging with FLAVORs is obtained by flow composition, FLAVORs have an intrinsic extension to multiscale structure preserving integrators for stiff Hamiltonian systems.

Recall that finite dimensional Hamiltonian systems in a Euclidean phase space

$\mathbb{R}^d \times \mathbb{R}^d$ could be described by ODEs of the form

$$\begin{cases} \dot{p} = -\partial_q \mathcal{H}(p,q) \\ \dot{q} = \partial_p \mathcal{H}(p,q) \end{cases} \tag{2.36}$$

where the most typical form of the Hamiltonian is

$$\mathcal{H}(q,p) := \frac{1}{2} p^T M^{-1} p + V(q), \tag{2.37}$$

which corresponds to the total energy of the system. Notice that the phase space does not have to be Euclidean, but could be as general as a cotangent bundle $T^*\mathcal{M}$ of a configuration manifold $\mathcal{M}$ (possibly infinite dimensional).

Recall an integrator for (2.36) in Euclidean space is symplectic, if and only if its one-step update map $\Phi : (q,p) \mapsto (\tilde{q}, \tilde{p})$ satisfies

$$(\mathrm{D}\Phi)^T \cdot J \cdot \mathrm{D}\Phi = J, \tag{2.38}$$

where

$$\mathrm{D}\Phi = \begin{bmatrix} \frac{\partial \tilde{q}}{\partial q}(q,p) & \frac{\partial \tilde{q}}{\partial p}(q,p) \\ \frac{\partial \tilde{p}}{\partial q}(q,p) & \frac{\partial \tilde{p}}{\partial p}(q,p) \end{bmatrix}$$

is the Jacobian matrix, and $J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$ corresponds to the symplectic 2-form.

Remark 2.2.1 (Variational integrator)**.** *One way to obtain a symplectic integrator is via a discrete variational principle, and the resulting integrator will be not only symplectic but also variational. The strategy is the following: assuming the corresponding Lagrangian $\mathcal{L}(q,\dot{q})$ is known (in the non-degenerate case it could be obtained as the Legendre transformation of $\mathcal{H}(q,p)$), then the exact dynamics (whose Hamiltonian form is given by* (2.36)*) is equivalently the critical point of the following action functional:*

$$\mathcal{S}[q(\cdot)] := \int_0^T \mathcal{L}(q(t), \dot{q}(t), t)\, dt \tag{2.39}$$

If one approximates the integral by a sum of discrete Lagrangians that satisfy

$$\mathcal{L}_d(q(kh), q((k+1)h)) \approx \int_{kh}^{(k+1)h} \mathcal{L}(q(t), \dot{q}(t), t)\, dt \tag{2.40}$$

for all k, where h is the integration timestep, then the critical point (with vanishing derivatives with respect to all discrete points q_k) of the discrete action

$$\mathcal{S}_d(q_0, q_1, \ldots, q_N) := \sum_{i=0}^{N-1} \mathcal{L}_d(q_i, q_{i+1}) \tag{2.41}$$

yields the discrete equations of motion (i.e., Discrete Euler-Lagrangian equations), which correspond to a variational (symplectic) integrator. More details of this construction could be found, for instance, in [192].

A simplest stiff Hamiltonian system corresponds to a sum of stiff and soft potentials, i.e., $\mathcal{H} = \frac{1}{2}p^T M^{-1} p + V(q) + \frac{1}{\epsilon}U(q)$, and the corresponding Hamilton's equations write as

$$\begin{cases} \dot{p} = -\nabla V(q) - \frac{1}{\epsilon}\nabla U(q) \\ \dot{q} = M^{-1} p \end{cases} \tag{2.42}$$

2.2.2 FLAVORs for Hamiltonian equations

Assume that we are given a first-order-accurate legacy integrator for (2.42) in which the parameter $1/\epsilon$ can be controlled, i.e., a mapping Φ_h^α acting on the phase space such that for $h \le h_0 \min(1, \alpha^{-\frac{1}{2}})$

$$\left| \Phi_h^\alpha(q,p) - (q,p) - h\big(M^{-1}p, -V(q) - \alpha U(q)\big) \right| \le Ch^2(1+\alpha) \tag{2.43}$$

Write Θ_δ, the FLAVOR discrete mapping approximating solutions of (2.42) over time steps $\delta \gg \epsilon$, i.e.,

$$(q_{(n+1)\delta}, p_{(n+1)\delta}) := \Theta_\delta(q_{n\delta}, p_{n\delta}). \tag{2.44}$$

FLAVOR can then be defined by

$$\Theta_\delta := \Phi^0_{\delta-\tau} \circ \Phi^{\frac{1}{\epsilon}}_{\tau} \tag{2.45}$$

Theorem 2.1.3 establishes the accuracy of this integrator under Conditions 2.1.5 and 2.1.6 provided that $\tau \ll \sqrt{\epsilon} \ll \delta$ and $\frac{\tau^2}{\epsilon} \ll \delta \ll \frac{\tau}{\sqrt{\epsilon}}$.

Remark 2.2.2. *We also refer to Remark 2.1.10 for the application of Theorem 2.1.3 to Hamiltonian systems. Consider for instance the linear Hamiltonian system* $H(q_1, q_2, p_1, p_2) := \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 + \frac{1}{2}q_1^2 + \frac{1}{\epsilon}(q_2 - q_1)^2$. *If the system is started from* $q_2(0) - q_1(0) = \mathcal{O}(\sqrt{\epsilon})$, *then the energy remains bounded as* $\epsilon \downarrow 0$ *and* (q_1, q_2, p_1, p_2) *F-converges due to the first part of Remark 2.1.10.*

For the same example, if the system is started from a point such that $q_2(0) - q_1(0) \neq \mathcal{O}(\sqrt{\epsilon})$ *then the energy in the system blows up as* $\epsilon \downarrow 0$, *the range of* $p_2 - p_1$ *blows up and it can therefore not converge, even in the sense of measures. However the (slow) process* $(q_1 + q_2, p_1 + p_2)$ *satisfies an equation of the type* (2.31) *where the dependence on fast variables is only through* $q_2 - q_1$ *(*$y^\alpha = q_2 - q_1$ *in* (2.31)*) and* $q_2 - q_1$ *is locally ergodic (as defined in Item 3 of Condition 2.1.5) and does converge in the sense of distributions. Henceforth if* $q_2(0) - q_1(0) \neq 0$ *then the generalization of Theorem 2.1.3 (see second part of Remark 2.1.10) applies with* η *non-injective.*

2.2.3 Structure preserving properties of FLAVORs

We will now show that FLAVORs inherit the structure preserving properties of their legacy integrators.

Theorem 2.2.1. *If for all* $h, \epsilon > 0$ Φ^ϵ_h *is symmetric under a group action, then* Θ_δ *is symmetric under the same group action.*

Theorem 2.2.2. *If* Φ^α_h *is symplectic on the co-tangent bundle* $T^*\mathcal{M}$ *of a configuration manifold* $\mathcal{M}$, *then* Θ_δ *defined by* (2.45) *is symplectic on the co-tangent bundle* $T^*\mathcal{M}$.

Theorem 2.2.1 and Theorem 2.2.2 can be resolved by noting that "the overall method is symplectic — as a composition of symplectic transformations, and it is symmetric — as a symmetric composition of symmetric steps" (see Chapter XIII.1.3 of [128]).

Write

$$\Phi_h^* := \left(\Phi_{-h}\right)^{-1} \tag{2.46}$$

Let us recall the following definition corresponding to Definition 1.4 of the Chapter V of [128]:

Definition 2.2.1. *A numerical one-step method Φ_h is called time-reversible if it satisfies $\Phi_h^* = \Phi_h$.*

The following theorem, whose proof is straightforward, shows how to derive a "symplectic and symmetric and time-reversible" FLAVOR from a symplectic legacy integrator and its adjoint. Since this derivation applies to manifolds, it also leads to structure-preserving FLAVORs for constrained mechanical systems.

Theorem 2.2.3. *If Φ_h^α is symplectic on the co-tangent bundle $T^*\mathcal{M}$ of a configuration manifold $\mathcal{M}$, then*

$$\Theta_\delta := \Phi_{\frac{\tau}{2}}^{\frac{1}{\epsilon},*} \circ \Phi_{\frac{\delta-\tau}{2}}^{0,*} \circ \Phi_{\frac{\delta-\tau}{2}}^{0} \circ \Phi_{\frac{\tau}{2}}^{\frac{1}{\epsilon}} \tag{2.47}$$

is symplectic and time-reversible on the co-tangent bundle $T^*\mathcal{M}$.

Remark 2.2.3. *Observe that (except for the first and last steps) iterating Θ_δ defined by* (2.47) *is equivalent to iterating*

$$\Theta_\delta := \Phi_{\frac{\delta-\tau}{2}}^{0,*} \circ \Phi_{\frac{\delta-\tau}{2}}^{0} \circ \Phi_{\frac{\tau}{2}}^{\frac{1}{\epsilon}} \circ \Phi_{\frac{\tau}{2}}^{\frac{1}{\epsilon},*} \tag{2.48}$$

It follows that a symplectic, symmetric and reversible FLAVOR can be obtained in a nonintrusive way from a Störmer/Verlet integrator for (2.42) *[131, 129, 292].*

2.2.4 An example of a symplectic FLAVOR

If the phase space is $\mathbb{R}^d \times \mathbb{R}^d$, then an example of symplectic FLAVOR is obtained from Theorem 2.2.2 by choosing Φ_h^α to be the symplectic Euler (also known as Variational Euler or VE for short) integrator defined by

$$\Phi_h^\alpha(q,p) = \begin{pmatrix} q \\ p \end{pmatrix} + h \begin{pmatrix} M^{-1}\Big(p - h\big(V(q) + \alpha U(q)\big)\Big) \\ -V(q) - \alpha U(q) \end{pmatrix} \tag{2.49}$$

and letting Θ_δ be defined by (2.45).

2.2.5 An example of a symplectic and time-reversible FLAVOR

If the phase space is the Euclidean space $\mathbb{R}^d \times \mathbb{R}^d$, then an example of symplectic and time-reversible FLAVOR is obtained by letting Θ_δ be defined by (2.47) with Φ_h^α being the symplectic Euler integrator given by (2.49) and its adjoint given by:

$$\Phi_h^{\alpha,*}(q,p) = \begin{pmatrix} q \\ p \end{pmatrix} + h \begin{pmatrix} M^{-1}p \\ -V(q + hM^{-1}p) - \alpha U(q + hM^{-1}p) \end{pmatrix} \tag{2.50}$$

2.2.6 An artificial FLAVOR

Natural FLAVORs defined by (2.27) (for instance, the nonintrusive FLAVOR given by (2.8)) are not the only ways to average the flows of (2.37). We present below an alternative method based on the freezing and unfreezing of degrees of freedom associated with stiff potentials. We have called this method 'artificial' because the legacy method cannot be used as a black box. In an example of this approach, the discrete flow approximating solutions of (2.42) is given by (2.44) with

$$\Theta_\delta := \theta^{tr}_{\delta-\tau} \circ \theta^{\epsilon}_{\tau} \circ \theta^{V}_{\delta} \tag{2.51}$$

where θ^V_δ is a symplectic map corresponding to the flow of $H^{slow}(q,p) := V(q)$, approximating the effects of the soft potential on momentum over the mesoscopic

time step δ and defined by

$$\theta_\delta^V\big(q,p\big) = \big(q, p - \delta\nabla V(q)\big). \tag{2.52}$$

θ_τ^ϵ is a symplectic map approximating the flow of $H^{fast}(q,p) := \frac{1}{2}p^T M^{-1}p + \frac{1}{\epsilon}U(q)$ over a microscopic time step τ:

$$\theta_\tau^\epsilon\big(q,p\big) = \big(q + \tau M^{-1}p, p - \frac{\tau}{\epsilon}\nabla U(q + tM^{-1}p)\big) \tag{2.53}$$

$\theta_{\delta-\tau}^{tr}$ is a map approximating the flow of the Hamiltonian $H^{free}(q,p) := \frac{1}{2}p^T M^{-1}p$ under a holonomic constraint imposing the freezing of the stiff potential U (i.e., in non-holonomic short-hand form, $\dot{U} = 0$). Velocities along the direction of constraints have to be stored and set to be 0 before the constrained dynamics, i.e., frozen, and the stored velocities should be restored after the constrained dynamics, i.e., unfrozen; geometrically speaking, one projects to the constrained subsymplectic manifold, runs the constrained dynamics, and lifts back to the original full space. Oftentimes, the exact solution to the constrained dynamics can be found (examples given in Section 2.5.3, 2.5.2, 2.6.2, 2.6.3 and 2.6.4).

When the exact solution to the constrained dynamics cannot be easily found, one may want to employ integrators for constrained dynamics such as SHAKE [236] or RATTLE [10] instead. This has to be done with caution, because symplecticity of the translational flow may be lost. The composition of projection onto the constrained manifold (freezing), evolution on the constrained manifold, and lifting from it to the unconstrained space (unfreezing) preserves symplecticity in the unconstrained space only if the evolution on the constrained manifold preserves the inherited symplectic form. A numerical integration preserves the discrete symplectic form on the constrained manifold, but not necessarily the projected continuous symplectic form.

Remark 2.2.4. *This artificial FLAVOR is locally a perturbation of nonintrusive*

FLAVORs. By splitting theory [199, 128],

$$\theta^{tr}_{\delta-\tau} \circ \theta^{\epsilon}_{\tau} \circ \theta^{V}_{\delta} \approx \theta^{tr}_{\delta-\tau} \circ \theta^{V}_{\delta-\tau} \circ \theta^{\epsilon}_{\tau} \circ \theta^{V}_{\tau} \approx \theta^{tr}_{\delta-\tau} \circ \theta^{V}_{\delta-\tau} \circ \Phi^{\frac{1}{\epsilon}}_{\tau} \tag{2.54}$$

whereas $\Phi^{0}_{\delta-\tau} \circ \Phi^{\frac{1}{\epsilon}}_{\tau} \approx \theta^{free}_{\delta-\tau} \circ \theta^{V}_{\delta-\tau} \circ \Phi^{\frac{1}{\epsilon}}_{\tau}$*, where* θ^{free} *is the flow of* $H^{free}(q,p)$ *under no constraint. The only difference is that constraints are treated in* θ^{tr} *but not in* θ^{free}*.*

Remark 2.2.5. *This artificial FLAVOR can be formally regarded as* $\Phi^{\infty}_{\delta-\tau} \circ \Phi^{\frac{1}{\epsilon}}_{\tau}$*. In contrast, the nonintrusive FLAVOR is* $\Phi^{0}_{\delta-\tau} \circ \Phi^{\frac{1}{\epsilon}}_{\tau}$*.*

The advantage of this artificial FLAVOR lies in the fact that only $\tau \ll \sqrt{\epsilon} \ll \delta$ and $\delta \ll \frac{\tau}{\sqrt{\epsilon}}$ are required for its accuracy (and not $\frac{\tau^2}{\epsilon} \ll \delta$). We also observe that, in general, artifical FLAVOR overperforms nonintrusive FLAVOR in FPU long time ($\mathcal{O}(\omega^2)$) simulations (we refer to Section 2.6.3).

2.2.7 Variational derivation of FLAVORs

FLAVORs based on variational legacy integrators [192] are variational too. Recall that discrete Lagrangian L_d is an approximation of the integral of the continuous Lagrangian over one time step, and the Discrete Euler-Lagrangian equation (DEL) is obtained by applying the variational (least action) principle to the discrete action, which is a sum of discrete Lagrangians. The following diagram commutes:

$$\begin{array}{ccc} Singlescale\ L_d & \xrightarrow{FLAVORization} & Multiscale\ L_d \\ \downarrow{\scriptstyle variational\ principle} & & \downarrow{\scriptstyle variational\ principle} \\ Singlescale\ DEL & \xrightarrow{FLAVORization} & Multiscale\ DEL \end{array}$$

For example, recall Variational Euler (i.e., symplectic Euler) for system (2.37) with time step h

$$\begin{cases} p_{k+1} &= p_k - h[\nabla V(q_k) + \frac{1}{\epsilon}\nabla U(q_k)] \\ q_{k+1} &= q_k + h p_{k+1} \end{cases} \tag{2.55}$$

can be obtained by applying variational principle to the following discrete La-

grangian

$$L_{d_h}^{1/\epsilon}(q_k, q_{k+1}) = h\left[\frac{1}{2}\left(\frac{q_{k+1}-q_k}{h}\right)^2 - \left(V(q_k) + \frac{1}{\epsilon}U(q_k)\right)\right]. \tag{2.56}$$

Meanwhile, FLAVORized Variational Euler with smallstep τ and mesostep δ

$$\begin{cases} p'_k & = p_k - \tau[\nabla V(q_k) + \frac{1}{\epsilon}\nabla U(q_k)] \\ q'_k & = q_k + \tau p'_k \\ p_{k+1} & = p'_k - (\delta-\tau)\nabla V(q'_k) \\ q_{k+1} & = q'_k + (\delta-\tau)p_{k+1} \end{cases} \tag{2.57}$$

can be obtained by applying variational principle to the FLAVORized discrete Lagrangian

$$\begin{aligned} &L_{d\delta}(q_k, q'_k, q_{k+1}) = L_{d_\tau}^{1/\epsilon}(q_k, q'_k) + L_{d_{\delta-\tau}}^{0}(q'_k, q_{k+1}) \\ &= \tau\left[\frac{1}{2}\left(\frac{q'_k-q_k}{\tau}\right)^2 - \left(V(q_k) + \frac{1}{\epsilon}U(q_k)\right)\right] + (\delta-\tau)\left[\frac{1}{2}\left(\frac{q_{k+1}-q'_k}{\delta-\tau}\right)^2 - V(q'_k)\right] \end{aligned} \tag{2.58}$$

FLAVORizations of other variational integrators such as Velocity Verlet follow similarly.

2.3 FLAVORs for general SDEs

2.3.1 Averaging

For the sake of clarity, we will start the description of with the following SDE on $\mathbb{R}^d$:

$$du_t^\epsilon = \left(G(u_t^\epsilon) + \frac{1}{\epsilon}F(u_t^\epsilon)\right)dt + \left(H(u_t^\epsilon) + \frac{1}{\sqrt{\epsilon}}K(u_t^\epsilon)\right)dW_t, \quad u_0^\epsilon = u_0 \tag{2.59}$$

where $(W_t)_{t\geq 0}$ is a d-dimensional Brownian motion; F and G are vector fields on $\mathbb{R}^d$; H and K are$d \times d$ matrix fields on $\mathbb{R}^d$. In Section 2.3.6, we will consider a more general form (2.73).

Regarding asymptotic problems for stochastic differential equations, we refer to Skorokhod's detailed monograph [258], as well as the early work of Gikhman [115], Krylov [169, 170], Bogolyubov [36] and Papanicolaou-Kohler [222]. Like in the ODE cases, effective equations for stiff SDEs can be obtained by averaging the instantaneous coefficients (drift and the diffusivity matrix squared) with respect to the fast components; Section 3 of Chapter II in [258] can be referred to for a detailed analysis including error bounds. Numerical methods such as HMM [87] and equation-free methods [15] have been extended to SDEs based on this averaging principle. Another idea is to treat fast variables by conditioning; here, we refer to optimal prediction [66, 65, 67] that has also been used for model reduction. Existing contributions also include [16, 126, 280, 52, 53, 184, 2].

In order for averaging to work, we again need conditions including a separation of timescales and a locally ergodic fast process:

Condition 2.3.1. *Assume that:*

1. *F, G, H and K are uniformly bounded and Lipschitz continuous.*

2. *There exists a diffeomorphism $\eta := (\eta^x, \eta^y)$, from $\mathbb{R}^d$ onto $\mathbb{R}^{d-p} \times \mathbb{R}^p$, independent of ϵ, with uniformly bounded C^1, C^2 and C^3 derivatives, such that the process $(x_t^\epsilon, y_t^\epsilon) = (\eta^x(u_t^\epsilon), \eta^y(u_t^\epsilon))$ satisfies the SDE*

$$\begin{cases} dx^\epsilon = g(x^\epsilon, y^\epsilon)\, dt + \sigma(x^\epsilon, y^\epsilon) dW_t, \quad x_0^\epsilon = x_0 \\ dy^\epsilon = \frac{1}{\epsilon} f(x^\epsilon, y^\epsilon)\, dt + \frac{1}{\sqrt{\epsilon}} Q(x^\epsilon, y^\epsilon) dW_t, \quad y_0^\epsilon = y_0 \end{cases} \tag{2.60}$$

   where g is $d-p$ dimensional vector field; f a p-dimensional vector field; σ is a $(d-p) \times d$-dimensional matrix field; Q a $p \times d$-dimensional matrix field and W_t a d-dimensional Brownian motion.

3. Let Y_t be the solution to

$$dY_t = f(x_0, Y_t)\,dt + Q(x_0, Y_t)\,dW_t, \qquad Y_0 = y_0, \tag{2.61}$$

there exists a family of probability measures $\mu(x, dy)$ on $\mathbb{R}^p$ indexed by $x \in \mathbb{R}^{d-p}$ and a positive function $T \mapsto E(T)$ such that $\lim_{T\to\infty} E(T) = 0$ and for all x_0, y_0, T and ϕ with uniformly bounded C^r derivatives for $r \leq 3$,

$$\Big|\frac{1}{T}\int_0^T \mathbb{E}\big[\phi(Y_s)\big] - \int \phi(y)\mu(x_0, dy)\Big| \leq \chi\big(\|(x_0, y_0)\|\big)E(T)\max_{r\leq 3}\|\phi\|_{C^r} \tag{2.62}$$

where $r \mapsto \chi(r)$ is bounded on compact sets.

4. For all u_0, $T > 0$, $\sup_{0\leq t\leq T}\mathbb{E}\Big[\chi\big(\|u_t^\epsilon\|\big)\Big]$ is uniformly bounded in ϵ.

Remark 2.3.1. *Like in the proof of Theorem 2.1.1, the uniform regularity of F, G, H and K can be relaxed to local regularity by adding a control on the rate of escape of the process towards infinity. To simplify the presentation, we will use the global uniform regularity.*

We will now extend the definition of two-scale flow convergence introduced in Section 2.1.3 to stochastic processes.

2.3.2 Two-scale flow convergence for SDEs

Let $\big(\xi_t^\epsilon(\omega)\big)_{t\in\mathbb{R}^+,\omega\in\Omega}$ be a sequence of stochastic processes on $\mathbb{R}^d$ (progressively measurable mappings from $\mathbb{R}^+ \times \Omega$ to $\mathbb{R}^d$) indexed by $\epsilon > 0$. Let $(X_t)_{t\in\mathbb{R}^+}$ be a (progressively measurable) stochastic process on $\mathbb{R}^{d-p}$ ($p \geq 0$). Let $x \mapsto \nu(x, dz)$ be a function from $\mathbb{R}^{d-p}$ into the space of probability measures on $\mathbb{R}^d$.

Definition 2.3.1. *We say that the process ξ_t^ϵ F-converges to $\nu(X_t, dz)$ as $\epsilon \downarrow 0$ and write $\xi_t^\epsilon \xrightarrow[\epsilon\to 0]{F} \nu(X_t, dz)$ if and only if for all function φ bounded and uniformly Lipschitz-continuous on $\mathbb{R}^d$ and all $t > 0$,*

$$\lim_{h\to 0}\lim_{\epsilon\to 0}\frac{1}{h}\int_t^{t+h}\mathbb{E}\big[\varphi(\xi_s^\epsilon)\big]\,ds = \mathbb{E}\big[\int_{\mathbb{R}^d}\varphi(z)\nu(X_t, dz)\big] \tag{2.63}$$

2.3.3 Nonintrusive FLAVORs for SDEs

Let ω be a random sample from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\Phi_h^\alpha(,\omega)$ a random mapping from $\mathbb{R}^d$ onto $\mathbb{R}^d$ approximating the flow of (2.59) with $\alpha = 1/\epsilon$. If the parameter α can be controlled, then Φ_h^α can be used as a black box for accelerating the computation of solutions of (2.59) without prior identification of slow variables. Indeed, assume that there exists a constant $h_0 > 0$ and a normal-distributed random vector $\xi(\omega)$ such that for $h \leq h_0 \min(\frac{1}{\alpha}, 1)$

$$\left(\mathbb{E}\Big[\big|\Phi_h^\alpha(u,\omega) - u - hG(u) - \alpha h F(u) - \sqrt{h}H(u)\xi(\omega) - \sqrt{\alpha h}K(u)\xi(\omega)\big|^2\Big]\right)^{\frac{1}{2}} \leq Ch^{\frac{3}{2}}(1+\alpha)^{\frac{3}{2}} \tag{2.64}$$

then FLAVOR can be defined as the algorithm simulating the stochastic process

$$\begin{cases} \bar{u}_0 = u_0 \\ \bar{u}_{(k+1)\delta} = \Phi_{\delta-\tau}^0(.,\omega_k') \circ \Phi_\tau^{\frac{1}{\epsilon}}(\bar{u}_{k\delta}, \omega_k) \quad . \\ \bar{u}_t = \bar{u}_{k\delta} \quad \text{for} \quad k\delta \leq t < (k+1)\delta \end{cases} \tag{2.65}$$

where ω_k, ω_k' are i.i.d. samples from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\delta \leq h_0$ and $\tau \in (0, \delta)$ such that $\tau \leq h_0\epsilon$. Theorem 2.3.1 establishes the asymptotic accuracy of FLAVOR for $\tau \ll \epsilon \ll \delta$ and

$$\big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}} \ll \delta \ll \frac{\tau}{\epsilon}. \tag{2.66}$$

Remark 2.3.2. *ω_k simulates the randomness of the increment of the Brownian motion between times δk and $\delta k + \tau$. ω_k' simulates the randomness of the increment of the Brownian motion between times $\delta k + \tau$ and $\delta(k+1)$. The independence of ω_k and ω_k' is reflection of the independence of the increments of a Brownian motion.*

2.3.4 Convergence theorem

Theorem 2.3.1. *Let u^ϵ be the solution to* (2.59) *and $\bar{u}_t$ defined by* (2.65)*. Assume that equation* (2.64) *and Condition 2.3.1 are satisfied, then*

- *u_t^ϵ F-converges towards $\eta^{-1} * \big(\delta_{X_t} \otimes \mu(X_t, dy)\big)$ as $\epsilon \downarrow 0$ where X_t is the solution to*

$$dX_t = \int g(X_t, y)\, \mu(X_t, dy)\, dt + \bar{\sigma}(X_t)\, dB_t \qquad X_0 = x_0 \tag{2.67}$$

where $\bar{\sigma}$ is a $(d-p) \times (d-p)$ matrix field defined by

$$\bar{\sigma}\bar{\sigma}^T(x) = \int \sigma\sigma^T(x, y)\, \mu(x, dy) \tag{2.68}$$

and B_t a $(d-p)$-dimensional Brownian motion.

- *$\bar{u}_t$ F-converges towards $\eta^{-1} * \big(\delta_{X_t} \otimes \mu(X_t, dy)\big)$ as $\epsilon \downarrow 0$, $\tau \leq \delta$, $\frac{\tau}{\epsilon} \downarrow 0$, $\frac{\delta\epsilon}{\tau} \downarrow 0$ and $\big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}} \frac{1}{\delta} \downarrow 0$.*

The proof of convergence of SDEs of type (2.60) is classical, and a comprehensive monograph can be found in Chapter II of [258]. A proof of (mean squared) convergence of HMM applied to (2.60) (separated slow and fast variables) with $\sigma = 0$ has been obtained in [87]. A proof of (mean squared) convergence of the Equation-Free Method applied to (2.60) with $\sigma \neq 0$ but independent of fast variables has been obtained in [118]. Theorem 2.3.1 proves the convergence in distribution of FLAVOR applied to SDE (2.59) with hidden slow and fast processes. One of the main difficulties of the proof of Theorem 2.3.1 lies in the fact that we are not assuming that the noise on (hidden) slow variables is null or independent from fast variables. Without this assumption, x_t^ϵ converges only weakly towards X_t, the convergence of u^ϵ can only be weak and techniques for strong convergence can not be used. The proof of Theorem 2.3.1 relies on a powerful result by Skorokhod (Theorem 1 of Chapter II of [258]) stating that the convergence in distribution of a sequence of stochastic processes is implied by the convergence of their generators. We refer to Section A.2 of the appendix for the detailed proof of Theorem 2.3.1.

2.3.5 Natural FLAVORs for SDEs

As for ODEs, it is not necessary to use legacy integrators to obtain FLAVORs for SDEs. More precisely, Theorem 2.3.1 remains valid if FLAVORs are defined to be algorithms simulating the discrete process

$$\begin{cases} \bar{u}_0 = u_0 \\ \bar{u}_{(k+1)\delta} = \theta^G_{\delta-\tau}(.,\omega'_k) \circ \theta^\epsilon_\tau(\bar{u}_{k\delta},\omega_k) \\ \bar{u}_t = \bar{u}_{k\delta} \quad \text{for} \quad k\delta \leq t < (k+1)\delta \end{cases} \tag{2.69}$$

where ω_k, ω'_k are i.i.d. samples from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and θ^ϵ_τ and $\theta^G_{\delta-\tau}$ are two random mappings from $\mathbb{R}^d$ onto $\mathbb{R}^d$ satisfying the following Condition 2.3.2, which implies that $\theta^\epsilon_\tau(.,\omega)$ approximates in distribution the flow of (2.59) over time steps $\tau \ll \epsilon$ and $\theta^G_h(.,\omega)$ approximates in distribution the flow of

$$dv^\epsilon_t = G(v^\epsilon_t)\,dt + H(v^\epsilon_t)\,dW_t \tag{2.70}$$

over time steps $h \ll 1$.

Condition 2.3.2. *Assume that:*

1. *There exists $h_0, C > 0$ and a d-dimensional centered Gaussian vector $\xi(\omega)$ with identity covariance matrix such that for $h \leq h_0$,*

$$\left(\mathbb{E}\Big[\big|\theta^G_h(u,\omega) - u - hG(u) - \sqrt{h}H(u)\xi(\omega)\big|^2\Big]\right)^{\frac{1}{2}} \leq Ch^{\frac{3}{2}} \tag{2.71}$$

2. *There exists $\tau_0, C > 0$ and a d-dimensional centered Gaussian vector $\xi(\omega)$ with identity covariance matrix such that for $\frac{\tau}{\epsilon} \leq \tau_0$,*

$$\left(\mathbb{E}\Big[\big|\theta^\epsilon_\tau(u,\omega) - u - \tau G(u) - \frac{\tau}{\epsilon}F(u) - \sqrt{\tau}H(u)\xi(\omega) - \sqrt{\frac{\tau}{\epsilon}}K(u)\xi(\omega)\big|^2\Big]\right)^{\frac{1}{2}} \leq C\big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}} \tag{2.72}$$

3. *For all u_0, $T > 0$, $\sup_{0\leq n\leq T/\delta} \mathbb{E}\Big[\chi\big(\|\bar{u}_{n\delta}\|\big)\Big]$ is uniformly bounded in ϵ, $0 <$*

$\delta \le h_0$, $\tau \le \min(\tau_0\epsilon, \delta)$, *where* $\bar{u}$ *is defined by* (2.69).

2.3.6 FLAVORs for generic stiff SDEs

FLAVORs for stochastic systems have a natural generalization to SDEs on $\mathbb{R}^d$ of the form

$$du^{\alpha,\epsilon} = F(u^{\alpha,\epsilon}, \alpha, \epsilon)\, dt + K(u^{\alpha,\epsilon}, \alpha, \epsilon)\, dW_t \tag{2.73}$$

where $(W_t)_{t\ge 0}$ is a d-dimensional Brownian motion, and F and K are Lipschitz continuous in u.

Condition 2.3.3. *Assume that:*

1. $\gamma \mapsto F(u, \alpha, \gamma)$ *and* $\gamma \mapsto K(u, \alpha, \gamma)$ *are uniformly continuous in the neighborhood of* 0.

2. *There exists a diffeomorphism* $\eta := (\eta^x, \eta^y)$, *from* $\mathbb{R}^d$ *onto* $\mathbb{R}^{d-p} \times \mathbb{R}^p$, *independent from* ϵ, α, *with uniformly bounded* C^1, C^2 *and* C^3 *derivatives, and such that the stochastic process* $(x_t^\alpha, y_t^\alpha) = (\eta^x(u_t^{\alpha,0}), \eta^y(u_t^{\alpha,0}))$ *satisfies for all* $\alpha \ge 1$ *the SDE*

$$dx^\alpha = g(x^\alpha, y^\alpha)\, dt + \sigma(x^\alpha, y^\alpha)\, dW_t \quad x_0^\alpha = x_0 \tag{2.74}$$

where g *is* $d-p$ *dimensional vector field,* σ *is a* $(d-p)\times d$*-dimensional matrix field, and* g *and* σ *are uniformly bounded and Lipschitz continuous in* x *and* y.

3. *There exists a family of probability measures* $\mu(x, dy)$ *on* $\mathbb{R}^p$ *such that for all* x_0, y_0 $\big((x_0, y_0) := \eta(u_0)\big)$, T *and* φ *with uniformly bounded* C^r *derivatives for* $r \le 3$,

$$\Big|\frac{1}{T}\int_0^T \mathbb{E}\big[\varphi(y_s^\alpha)\big]\, ds - \int \varphi(y)\mu(x_0, dy)\Big| \le \chi\big(\|(x_0, y_0)\|\big)\big(E_1(T) + E_2(T\alpha^\nu)\big) \max_{r\le 3} \|\varphi\|_{C^r} \tag{2.75}$$

where $r \mapsto \chi(r)$ is bounded on compact sets and $E_2(r) \to 0$ as $r \to \infty$ and $E_1(r) \to 0$ as $r \to 0$.

4. *For all u_0, $T > 0$, $\sup_{0 \leq t \leq T} \mathbb{E}\Big[\chi\big(\|u_t^{\alpha,0}\|\big)\Big]$ is uniformly bounded in $\alpha \geq 1$.*

Remark 2.3.3. *Like in the proof of Theorem 2.1.1, the uniform regularity of g and σ can be relaxed to local regularity by adding a control on the rate of escape of the process towards infinity. To simplify the presentation, we have used the global uniform regularity.*

Let ω be a random sample from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\Phi_h^{\alpha,\epsilon}(.,\omega)$ a random mapping from $\mathbb{R}^d$ onto $\mathbb{R}^d$ approximating in distribution the flow of (2.73) over time steps $\tau \ll \epsilon$. If the parameter α can be controlled, then $\Phi_h^{\alpha,\epsilon}$ can be used as a black box for accelerating the computation of solutions of (2.73). The acceleration is obtained without prior identification of the slow variables.

Condition 2.3.4. *Assume that:*

1. *There exists $h_0, C, \nu > 0$ and a d-dimensional centered Gaussian vector $\xi(\omega)$ with identity covariance matrix such that for $h \leq h_0$, $0 < \epsilon \leq 1 \leq \alpha$ and $h \leq h_0 \min(\frac{1}{\alpha^\nu}, 1)$*

$$\left(\mathbb{E}\Big[\big|\Phi_h^{\alpha,\epsilon}(u) - u - hF(u,\alpha,\epsilon) - \sqrt{h}\xi(\omega)K(u,\alpha,\epsilon)\big|^2\Big)^{\frac{1}{2}} \leq Ch^{\frac{3}{2}}(1+\alpha^{\frac{3\nu}{2}}) \right. \tag{2.76}$$

2. *For all u_0, $T > 0$, $\sup_{0 \leq n \leq T/\delta} \mathbb{E}\Big[\chi\big(\|\bar{u}_{n\delta}\|\big)\Big]$ is uniformly bounded in ϵ, $0 < \delta \leq h_0$, $\tau \leq \min(h_0\epsilon^\nu, \delta)$, where $\bar{u}$ is defined by (2.77).*

Let $\delta \leq h_0$ and $\tau \in (0, \delta)$ such that $\tau \leq \tau_0 \epsilon^\nu$. Then a FLAVOR integration is defined as the stochastic process $t \mapsto \bar{u}_t$ given by

$$\begin{cases} \bar{u}_0 = u_0 \\ \bar{u}_{(k+1)\delta} = \Phi_{\delta-\tau}^{0,\epsilon}(.,\omega_k') \circ \Phi_\tau^{\frac{1}{\epsilon},\epsilon}(\bar{u}_{k\delta}, \omega_k) \\ \bar{u}_t = \bar{u}_{k\delta} \quad \text{for} \quad k\delta \leq t < (k+1)\delta \end{cases} \tag{2.77}$$

where ω_k, ω_k' are i.i.d. samples from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

The following theorem shows that the flow averaging integrator is accurate with respect to F-convergence for $\tau \ll \epsilon^\nu \ll \delta$ and

$$\left(\frac{\tau}{\epsilon^\nu}\right)^{\frac{3}{2}} \ll \delta \ll \frac{\tau}{\epsilon^\nu}. \tag{2.78}$$

Theorem 2.3.2. *Let $u_t^{\frac{1}{\epsilon},\epsilon}$ be the solution to (2.73) with $\alpha = 1/\epsilon$ and $\bar{u}_t$ be defined by (2.77). Assume that Conditions 2.3.3 and 2.3.4 are satisfied then*

- *$u_t^{\frac{1}{\epsilon},\epsilon}$ F-converges towards $\eta^{-1} * \left(\delta_{X_t} \otimes \mu(X_t, dy)\right)$ as $\epsilon \downarrow 0$ where X_t is the solution to*

$$dX_t = \int g(X_t, y)\, \mu(X_t, dy) + \bar{\sigma}(X_t)\, dB_t \qquad X_0 = x_0 \tag{2.79}$$

  where $\bar{\sigma}$ is a $(d-p) \times (d-p)$ matrix field defined by

$$\bar{\sigma}\bar{\sigma}^T(x) = \int \sigma\sigma^T(x, y)\, \mu(x, dy) \tag{2.80}$$

  and B_t a $(d-p)$-dimensional Brownian motion.

- *As $\epsilon \downarrow 0$, $\tau\epsilon^{-\nu} \downarrow 0$, $\delta\frac{\epsilon^\nu}{\tau} \downarrow 0$, $\left(\frac{\tau}{\epsilon^\nu}\right)^{\frac{3}{2}}\frac{1}{\delta} \downarrow 0$, the numerical solution $\bar{u}_t$ F-converges towards $\eta^{-1} * \left(\delta_{X_t} \otimes \mu(X_t, dy)\right)$ as $\epsilon \downarrow 0$ where X_t is the solution to (2.79).*

Proof. The proof of Theorem 2.3.2 is similar to the proof of Theorem 2.3.1. The condition $\epsilon \ll 1$ is needed for the approximation of $u^{\alpha,\epsilon}$ by $u^{\alpha,0}$ and for the F-convergence of $u^{\frac{1}{\epsilon},0}$. Since $y_t^\alpha = \eta^y(u_t^{\alpha,0})$ the condition $\tau \ll \epsilon^\nu$ is used along with (2.76) for the accuracy of $\Phi_\tau^{\frac{1}{\epsilon},\epsilon}$ in (locally) approximating y_t^α. The condition $\delta \ll \frac{\tau}{\epsilon^\nu}$ allows for the averaging of g and σ to take place prior to a significant change of $x\alpha_t$; more precisely, it allows for $m \gg 1$ iterations of $\Phi_\tau^{\frac{1}{\epsilon},\epsilon}$ prior to a significant change of $x\alpha_t$. The condition $\left(\frac{\tau}{\epsilon^\nu}\right)^{\frac{3}{2}} \ll \delta$ is required in order to control the error accumulated by m iterations of $\Phi_\tau^{\frac{1}{\epsilon},\epsilon}$. □

2.4 FLAVORs for stochastic mechanical systems

2.4.1 Langevin system and Boltzmann-Gibbs distribution

Stiff stochastic mechanical systems are commonly modeled by stiff Langevin equations of the form

$$\begin{cases} dq = M^{-1}p \\ dp = -\nabla V(q)\,dt - \frac{1}{\epsilon}\nabla U(q)\,dt - cp\,dt + \sqrt{2\beta^{-1}}c^{\frac{1}{2}}dW_t \end{cases} \tag{2.81}$$

or

$$\begin{cases} dq = M^{-1}p \\ dp = -\nabla V(q)\,dt - \frac{1}{\epsilon}\nabla U(q)\,dt - \frac{c}{\epsilon}p\,dt + \sqrt{2\beta^{-1}}\frac{c^{\frac{1}{2}}}{\sqrt{\epsilon}}dW_t, \end{cases} \tag{2.82}$$

where c is a positive symmetric $d \times d$ matrix indicating the strength of the viscous friction, and $\beta \geq 0$ is a real number, also known as the inverse of the temperature.

(2.81) and (2.82) model a stochastic mechanical system whose deterministic part is given by a Hamiltonian

$$H(q,p) := \frac{1}{2}p^T M^{-1}p + V(q) + \frac{1}{\epsilon}U(q), \tag{2.83}$$

and additional internal dissipation due to friction and external perturbation due to noise are included. The phase space is the Euclidean space $\mathbb{R}^d \times \mathbb{R}^d$, but again it could be as general as a cotangent bundle $T^*\mathcal{M}$ of a configuration manifold $\mathcal{M}$.

Remark 2.4.1. *In* (2.81)*, both the dissipation and the noise are weak (soft), whereas they are both strong (stiff) in* (2.82)*. Provided that hidden fast variables remain locally ergodic, one can also consider a mixture of both soft and stiff noise and friction. For the sake of clarity, we have restricted our presentation to* (2.81) *and* (2.82)*.*

Remark 2.4.2. *If c is not constant and $\mathcal{M}$ is not the usual $\mathbb{R}^d \times \mathbb{R}^d$ Euclidean space, one should use the Stratonovich integral instead of the Itô integral.*

The energy of the system (given by the Hamiltonian function) will no longer be

conserved, but fluctuating according to a balance between the injection of energy due the noise and the loss of energy due to the dissipation. Naturally, the amplitude of fluctuation positively scales with the temperature $1/\beta$ (in the statistical sense). More precisely, define a probability distribution of Boltzmann-Gibbs on the phase space by

$$d\mu = Z^{-1} \exp(-\beta H(q,p)) dq dp, \tag{2.84}$$

where $Z = \int_{T^*\mathbb{R}^d} \exp(-\beta H(q,p))\, dq\, dp$ is the partition function, and H is the Hamiltonian. Then in many cases (see for instance [198] for sufficient conditions) Boltzmann-Gibbs is the ergodic and invariant measure of (2.81) or (2.82).

2.4.2 FLAVORs for Langevin equations

Like the case in Section 2.2, we assume that we are given a mapping Φ_h^α acting on the phase space such that for $h \le h_0 \min(1, \alpha^{-\frac{1}{2}})$

$$\left| \Phi_h^\alpha(q,p) - (q,p) - h\big(M^{-1}p, -V(q) - \alpha U(q)\big) \right| \le Ch^2(1+\alpha) \tag{2.85}$$

Next, consider the following Ornstein-Uhlenbeck equations:

$$dp = -\alpha c p\, dt + \sqrt{\alpha}\sqrt{2\beta^{-1}} c^{\frac{1}{2}} dW_t \tag{2.86}$$

The stochastic flow of (2.86) is defined by the following stochastic evolution map:

$$\Psi_{t_1,t_2}^\alpha(q,p) = \Big(q, e^{-c\alpha(t_2-t_1)}p + \sqrt{2\beta^{-1}\alpha} c^{\frac{1}{2}} \int_{t_1}^{t_2} e^{-c\alpha(t_2-s)} dW_s\Big) \tag{2.87}$$

Let $\delta \le h_0$ and $\tau \in (0,\delta)$ such that $\tau \le \tau_0/\sqrt{\alpha}$. FLAVOR for (2.81) can then be defined by

$$\begin{cases} (\bar{q}_0, \bar{p}_0) = (q_0, p_0) \\ (\bar{q}_{(k+1)\delta}, \bar{p}_{(k+1)\delta}) = \Phi_{\delta-\tau}^0 \circ \Psi_{k\delta+\tau,(k+1)\delta}^1 \circ \Phi_\tau^{\frac{1}{\epsilon}} \circ \Psi_{k\delta,k\delta+\tau}^1(\bar{q}_{k\delta}, \bar{p}_{k\delta}) \end{cases} \tag{2.88}$$

and FLAVOR for (2.82) can be defined by

$$\begin{cases} (\bar{q}_0, \bar{p}_0) = (q_0, p_0) \\ (\bar{q}_{(k+1)\delta}, \bar{p}_{(k+1)\delta}) = \Phi^0_{\delta-\tau} \circ \Phi^{\frac{1}{\epsilon}}_{\tau} \circ \Psi^{\frac{1}{\epsilon}}_{k\delta,k\delta+\tau}(\bar{q}_{k\delta}, \bar{p}_{k\delta}) \end{cases} \tag{2.89}$$

Theorem 2.3.2 establishes the accuracy of these integrators under Conditions 2.3.3 and 2.3.4 provided that $\tau \ll \sqrt{\epsilon} \ll \delta$ and $\left(\frac{\tau}{\sqrt{\epsilon}}\right)^{\frac{3}{2}} \ll \delta \ll \frac{\tau}{\sqrt{\epsilon}}$.

Remark 2.4.3. *Notice that a single-scale Geometric Langevin Integrator [41] can be constructed by using a one-step update map $\theta_{kh,(k+1)h} = \Phi^{\alpha}_{h} \circ \Psi^{\alpha}_{kh,(k+1)h}$. In this sense, the above is just a FLAVORization of θ.*

2.4.3 Structure preserving properties of FLAVORs

First, observe that if Φ^{α}_{h} and $\Psi^{\frac{1}{\epsilon}}_{h}$ are symmetric under a group action for all $\epsilon > 0$, then the resulting FLAVOR, as a symmetric composition of symmetric steps, is symmetric under the same group action (analogous to Theorem 2.2.1).

Similarly, the following theorem shows that FLAVORs inherit structure-preserving properties from those associated with Φ^{α}_{h} (the component approximating the Hamiltonian part of the flow).

Theorem 2.4.1.

- *If Φ^{α}_{h} is symplectic, then the FLAVORs defined by* (2.88) *and* (2.89) *are quasi-symplectic as defined in Conditions RL1 and RL2 of [205] (it degenerates to a symplectic method if friction is set equal to zero and the Jacobian of the flow map is independent of (q, p)).*

- *If in addition c is isotropic then FLAVOR defined by* (2.88) *is conformally symplectic, i.e., it preserves the precise symplectic area change associated to the flow of inertial Langevin processes [202].*

Proof. Those properties are a consequence of the fact that FLAVORs are composition schemes. The quasi-symplecticity and conformal-symplecticity of GLA [41] has been obtained in a similar way. □

Remark 2.4.4. *Quasi-symplecticity and conformal-symplecticity highly correlate to the convergence towards and the preservation of a near-by Boltzmann-Gibbs in the single-scale case (we refer to the proof in [41]). In the multiscale case considered here, the Boltzmann-Gibbs corresponding to the FLAVOR solution (if exists) will by no means be near by in the usual sense (i.e., in total variation norm), but it may be given by an effective Hamiltonian that no longer contains the stiffness.*

2.4.4 An example of a quasi-symplectic FLAVOR

An example of quasi-symplectic FLAVOR can be obtained by choosing Φ_h^α to be the symplectic Euler integrator defined by (2.49) or (2.50). This integrator is also conformally symplectic if c is isotropic and soft (with $\mathcal{O}(1)$ norm).

2.4.5 An example of a quasi-symplectic and time-reversible FLAVOR

Defining Φ_h^α by (2.49) and $\Phi_h^{\alpha,*}$ by (2.50), an example of quasi-symplectic and time-reversible FLAVOR can be obtained by using the symmetric Strang splitting:

$$(\bar{q}_{(k+1)\delta}, \bar{p}_{(k+1)\delta}) = \Psi^1_{k\delta+\frac{\delta}{2},(k+1)\delta} \circ \Phi^{\frac{1}{\epsilon}\frac{\tau}{2},*} \circ \Phi^{0,*}_{\frac{\delta-\tau}{2}} \circ \Phi^0_{\frac{\delta-\tau}{2}} \circ \Phi^{\frac{1}{\epsilon}\frac{\tau}{2}} \circ \Psi^1_{k\delta,k\delta+\frac{\delta}{2}}(q,p) \quad (2.90)$$

for (2.81) and

$$(\bar{q}_{(k+1)\delta}, \bar{p}_{(k+1)\delta}) = \Psi^{\frac{1}{\epsilon}}_{(k+1)\delta-\frac{\tau}{2},(k+1)\delta} \circ \Phi^{\frac{1}{\epsilon}\frac{\tau}{2},*} \circ \Phi^{0,*}_{\frac{\delta-\tau}{2}} \circ \Phi^0_{\frac{\delta-\tau}{2}} \circ \Phi^{\frac{1}{\epsilon}\frac{\tau}{2}} \circ \Psi^{\frac{1}{\epsilon}}_{k\delta,k\delta+\frac{\tau}{2}}(q,p) \quad (2.91)$$

for (2.82). Notice the symmetrization of (2.88) will be

$$\begin{aligned}(\bar{q}_{(k+1)\delta}, \bar{p}_{(k+1)\delta}) = \Psi^1_{(k+1)\delta-\frac{\tau}{2},(k+1)\delta} \circ \Phi^{\frac{1}{\epsilon}\frac{\tau}{2},*} \circ \Psi^1_{k\delta+\frac{\delta}{2},(k+1)\delta-\frac{\tau}{2}} \circ \Phi^{0,*}_{\frac{\delta-\tau}{2}} \\ \circ \Phi^0_{\frac{\delta-\tau}{2}} \circ \Psi_{k\delta+\frac{\tau}{2},k\delta+\frac{\delta}{2}} \circ \Phi^{\frac{1}{\epsilon}\frac{\tau}{2}} \circ \Psi^1_{k\delta,k\delta+\frac{\tau}{2}}(q,p), \quad (2.92)\end{aligned}$$

which is slightly more complicated than (2.90), but in fact both of them work (they are locally equivalent due to splitting theory).

These integrators are also conformally symplectic if c is isotropic and soft (with $\mathcal{O}(1)$ norm).

2.4.6 An example of Boltzmann-Gibbs reversible Metropolis-adjusted FLAVOR

Geometric Langevin Algorithm [41] is not stochastically stable if the vector field that it integrates is only locally but not globally Lipschitz. The statistics community has been composing a Metropolis step (which includes a proposed local momentum flip) with a usual integration step to overcome this difficulty and obtain ergodicity (if the noise applied on momentum is not degenerate), and the convergence towards a near-by Boltzmann-Gibbs by doing so is proved [42].

This Metropolis-Adjusted Geometric Langevin Algorithm can also be FLAVORized: since the probability density of Ψ_{t_1,t_2} can be explicitly computed, it follows that the probability densities of (2.90) and (2.91) can be explicitly computed, and therefore these algorithms can be Metropolized and made reversible with respect to the Boltzmann-Gibbs distribution. This Metropolization leads to stochastical stability, as well as ergodicity if the noise applied on momentum is not degenerate. Observe that if the proposed move is rejected, the momentum has to be flipped and the acceptance probability involves a momentum flip. It is proven in [42] that GLA [41] remains strongly accurate after a Metropolization involving local momentum flips. Whether this preservation of accuracy over trajectories transfers in a weak sense (in distributions) to FLAVORs remains to be investigated.

2.5 Numerical analysis of FLAVOR based on Variational Euler

2.5.1 Stability

Consider the following linear Hamiltonian system

$$H(x, y, p_x, p_y) = \frac{1}{2}p_x^2 + \frac{1}{2}p_y^2 + \frac{1}{2}x^2 + \frac{\omega^2}{2}(y - x)^2 \tag{2.93}$$

with $\omega \gg 1$. Here $\frac{x+y}{2}$ is the slow variable and $y - x$ is the fast variable.

It can be shown that, when applied to (2.93), Variational Euler (also known as symplectic Euler, i.e., (2.49)) is stable if and only if $h \leq \sqrt{2}/\omega$. Write $\Theta_{\delta,\tau}$ the non-intrusive FLAVOR (2.45) obtained by using Symplectic Euler (2.49) as the legacy integrator. Write $\Theta^a_{\delta,\tau}$ the artificial FLAVOR described in Section 2.2.6.

Theorem 2.5.1. *The non-intrusive FLAVOR $\Theta_{\delta,\tau}$ with $1/\sqrt{\tau} \gg \omega \gg 1$ is stable if and only if $\delta \in (0, 2)$.*
The artificial FLAVOR $\Theta^a_{\delta,\tau}$ with $1/\tau \gg \omega \gg 1$ is stable if and only if $\delta \in (0, 2\sqrt{2})$.

Proof. The numerical scheme associated with $\Theta_{\delta,\tau}$ can be written as

$$\begin{bmatrix} y_{n+1} \\ x_{n+1} \\ (p_y)_{n+1} \\ (p_x)_{n+1} \end{bmatrix} = T \begin{bmatrix} y_n \\ x_n \\ (p_y)_n \\ (p_x)_n \end{bmatrix} \tag{2.94}$$

with

$$T = \begin{bmatrix} 1 & 0 & \delta - \tau & 0 \\ 0 & 1 & 0 & \delta - \tau \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \tau - \delta & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & \tau & 0 \\ 0 & 1 & 0 & \tau \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\tau(\omega^2 + 1) & \tau\omega^2 & 1 & 0 \\ \tau\omega^2 & -\tau\omega^2 & 0 & 1 \end{bmatrix}$$

The characteristic polynomial of T is

$$\begin{aligned}&\lambda^4+(-4+\delta^2-\delta^2\tau^2+2\delta\tau^3-\tau^4+2\delta\tau\omega^2-\delta^2\tau^2\omega^2+2\delta\tau^3\omega^2-\tau^4\omega^2)\lambda^3+(6-2\delta^2\\&+2\delta^2\tau^2-4\delta\tau^3+2\tau^4-4\delta\tau\omega^2+\delta^3\tau\omega^2+2\delta^2\tau^2\omega^2-4\delta\tau^3\omega^2-\delta^3\tau^3\omega^2+2\tau^4\omega^2\\&+2\delta^2\tau^4\omega^2-\delta\tau^5\omega^2)\lambda^2\\&+(-4+\delta^2-\delta^2\tau^2+2\delta\tau^3-\tau^4+2\delta\tau\omega^2-\delta^2\tau^2\omega^2+2\delta\tau^3\omega^2-\tau^4\omega^2)\lambda+1\end{aligned}\tag{2.95}$$

Since $\omega\gg 1$, $\tau\ll 1/\omega^2$, as long as $\delta\lesssim 1$, roots to the above polynomial are (by continuity; we refer for instance to Theorem 1 of [78]) close to roots to the asymptotic polynomial

$$\lambda^4+(\delta^2-4)\lambda^3+(6-2\delta^2)\lambda^2+(\delta^2-4)\lambda+1\tag{2.96}$$

which can be shown to be 1 with multiplicity 2 and $\frac{1}{2}(2-\delta^2\pm\delta\sqrt{\delta^2-4})$. It is easy to see that all roots are complex numbers with moduli less or equal to one if and only if $|\delta|\leq 2$.

The numerical scheme associated with $\Theta^a_{\delta,\tau}$ can be written as in (2.94) with

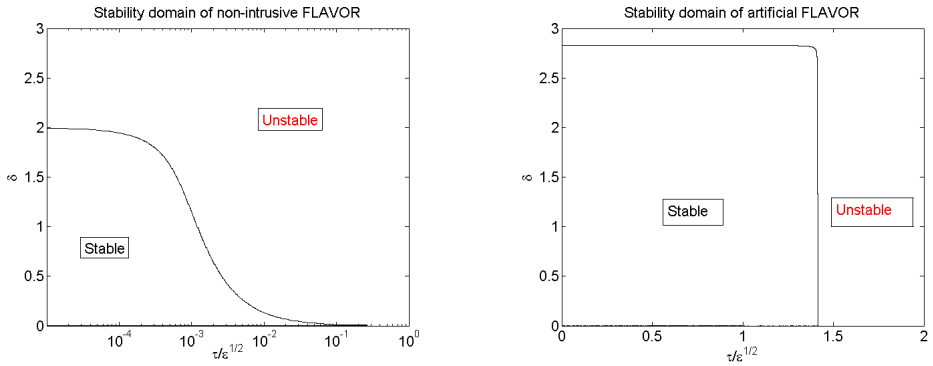
$$T=\begin{bmatrix}1&0&\frac{\delta-\tau}{2}&\frac{\delta-\tau}{2}\\0&1&\frac{\delta-\tau}{2}&\frac{\delta-\tau}{2}\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&0&0\\-\tau\omega^2&\tau\omega^2&1&0\\\tau\omega^2&-\tau\omega^2&0&1\end{bmatrix}\begin{bmatrix}1&0&\tau&0\\0&1&0&\tau\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&0&0\\-\delta&0&1&0\\0&0&0&1\end{bmatrix}\tag{2.97}$$

The characteristic polynomial of T is

$$2\lambda^4+(4\omega^2\tau^2+\tau\delta+\delta^2-8)\lambda^3+(12-2\delta^2-2\delta\tau-8\tau^2\omega^2+2\delta^2\tau^2\omega^2)\lambda^2+(4\omega^2\tau^2+\tau\delta+\delta^2-8)\lambda+2\tag{2.98}$$

Similarly, since $\omega\gg 1$, $\tau\ll 1/\omega$, as long as $\delta\lesssim 1$, roots to the above polynomial are close to roots to the asymptotic polynomial

$$2\lambda^4+(\delta^2-8)\lambda^3+(12-2\delta^2)\lambda^2+(\delta^2-8)\lambda+1\tag{2.99}$$



(a) *Nonintrusive FLAVOR* **(b) *Artificial FLAVOR***

Figure 2.2: Stability domain of non-intrusive and artificial FLAVOR applied to (2.93) as a function of δ and τ/ϵ. $\omega = 1/\sqrt{\epsilon} = 1000$.

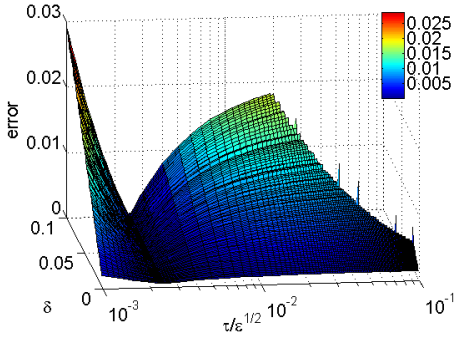
which can be shown to be 1 with multiplicity 2 and $\frac{1}{4}(4-\delta^2 \pm \delta\sqrt{\delta^2-8})$. All roots are complex numbers with moduli less or equal to one if and only if $|\delta| \leq 2\sqrt{2}$ $\square$

Figures 2.2(a) and 2.2(b) illustrate the domain of stability of nonintrusive FLAVOR (based on symplectic Euler (2.45) and (2.49)) and artificial FLAVOR (2.51) applied to the flow of (2.93), i.e., values of δ and τ/ϵ ensuring stable numerical integrations. We observe that artificial FLAVOR has a much larger stability domain than nonintrusive FLAVOR. Specifically, for nonintrusive FLAVOR and large values of δ, $\tau = o(\sqrt{\epsilon})$ is not enough and one needs $\tau = o(\epsilon)$ for a stable integration, whereas artificial FLAVOR only requires $\tau = \sqrt{2\epsilon}$, a minimum requirement for a stable symplectic Euler integration of the fast dynamics.

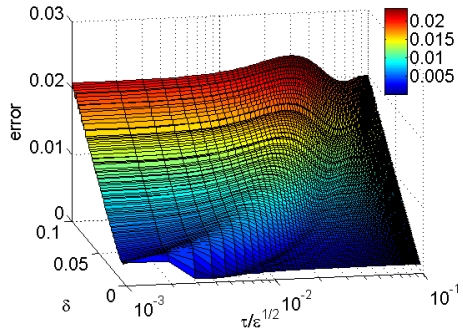
Notice that there is no resonance behavior in terms of stability; everything below the two curves is stable and everything outside is not stable (plots not shown).

2.5.2 Error analysis

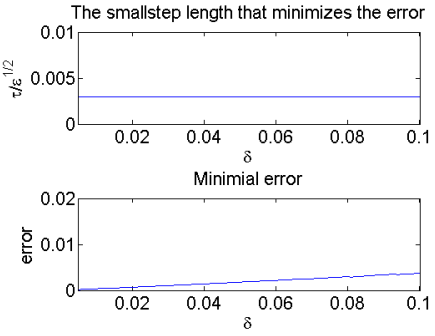
The flow of (2.93) has been explicitly computed and compared with solutions obtained from nonintrusive FLAVOR based on symplectic Euler ((2.45) and (2.49)) and with artificial FLAVOR (2.51).



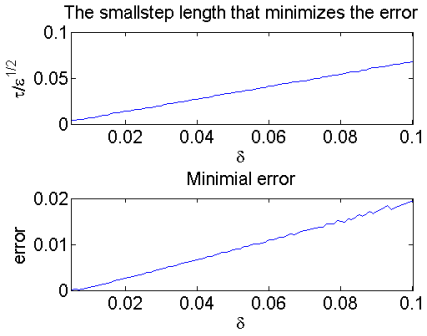
(a) ***Error of nonintrusive FLAVOR as a function of δ and $\tau/\sqrt{\epsilon}$. Notice that not all pairs of step lengths lead to stable integrations.***



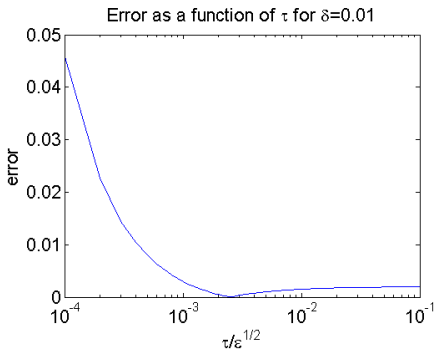
(b) ***Error of artificial FLAVOR as a function of δ and $\tau/\sqrt{\epsilon}$***



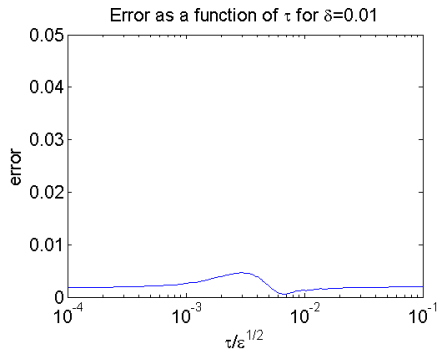
(c) ***Optimal $\tau/\sqrt{\epsilon}$ and error of nonintrusive FLAVOR as functions of δ***



(d) ***Optimal $\tau/\sqrt{\epsilon}$ and error of artificial FLAVOR as functions of δ***



(e) ***Error dependence on $\tau/\sqrt{\epsilon}$ for a given δ: nonintrusive FLAVOR***



(f) ***Error dependence on $\tau/\sqrt{\epsilon}$ for a given δ: artificial FLAVOR***

Figure 2.3: Error analysis of (2.93). Parameters are $\omega = \sqrt{\epsilon} = 10^3$, $x(0) = 0.8$ and $y(0) = x(0) + 1.1/\omega$.

The total simulation time is $T = 10$, and absolute errors on the slow variable have been computed using the Euclidean distance between the final positions of the analytical solution and the numerical solution. Figures 2.3(a) and 2.3(b) illustrate errors as functions of the mesostep δ and the scaled microstep τ/ϵ. Observe that given δ, errors are minimized at specific values of τ/ϵ for both integrators, but the accuracy of nonintrusive FLAVOR is less sensitive to τ/ϵ. Figures 2.3(c) and 2.3(d) plot the optimal value of τ/ϵ as a function of δ and the associated error. Observe also that for nonintrusive FLAVOR, the optimal value of τ/ϵ only weakly depends on δ, whereas for artificial FLAVOR the optimal value of τ/ϵ roughly scales linearly with δ. Figure 2.3(e) and 2.3(f) describe how error changes with microstep τ while mesostep δ is fixed. Figure 2.3(e) can be viewed in correspondence with the condition $\delta \ll \tau/\epsilon$ required for accuracy. This requirement, however, is just a sufficient but not necessary condition to obtain an error bound, as we can see in Figure 2.3(f). There, the weak dependence of the error on τ/ϵ (for a fixed δ) shows that one does not have to choose the microstep with too much care nor optimize the integrator with respect to its value, if an artificial FLAVOR is used. As a matter of fact, all the numerical experiments illustrated in this chapter (except for Figures 2.3(c) and 2.3(d)) have been performed without any tuning of the τ/ϵ value. We have simply used the rule of thumb $\delta \sim \gamma\frac{\tau}{\epsilon}$ where γ is a small parameter (0.1 for instance).

Final remarks are, (i) there is no resonant value of δ or τ, and (ii) it appears that the benefits of artificial FLAVORs lie in their superior accuracy and stability.

2.5.3 Numerical error analysis on a nonlinear system

In this section, we will consider the nonlinear Hamiltonian system

$$H(x,y,z,p_x,p_y,p_z) = \frac{1}{2}p_x^2+\frac{1}{2}p_y^2+\frac{1}{2}p_z^2+x^4+\epsilon^{-1}\frac{\omega_1}{2}(y-x)^2+\epsilon^{-1}\frac{\omega_2}{2}(z-y)^2 \quad (2.100)$$

That is, the stiff potential is $U = \frac{\omega_1}{2}(y-x)^2 + \frac{\omega_2}{2}(z-y)^2$, and the soft potential is $V = x^4$. Here $\frac{x+y+z}{3}$ acts as a slow degree of freedom and $y-x$ and $z-y$ act

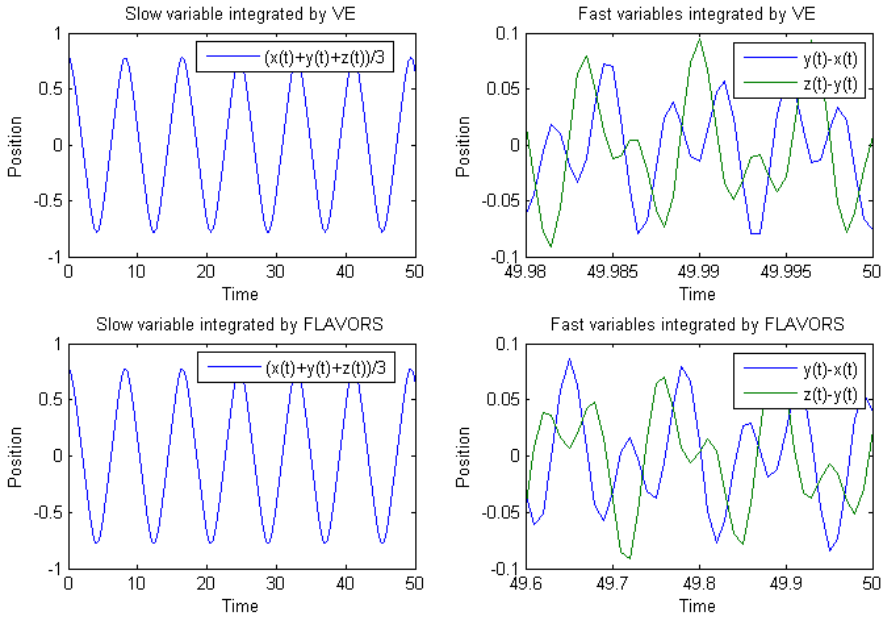
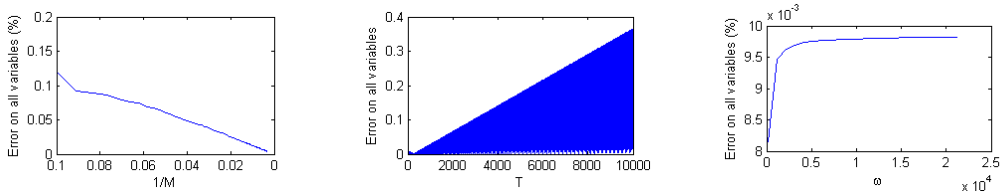


Figure 2.4: Comparison between trajectories integrated by Variational Euler and FLAVOR (defined by (2.45) and (2.49)). FLAVOR uses mesostep $\delta = 0.01$ and microstep $\tau = 0.0005$, and Symplectic Euler uses time step $\tau = 0.0005$. Time axes in the right column are zoomed in (by different ratios) to illustrate the fact that fast variables are captured in the sense of measure. FLAVOR accelerated the computation by roughly 20x ($\delta = 20\tau$). In this experiment $\epsilon = 10^{-6}$, $\omega_1 = 1.1$, $\omega_2 = 0.97$, $x(0) = 0.8$, $y(0) = 0.811$, $z(0) = 0.721$, $p_x(0) = 0$, $p_y(0) = 0$ and $p_z(0) = 0$. Simulation time $T = 50$.

as fast degrees of freedom.

Figure 2.4 illustrates $t \mapsto \frac{x(t)+y(t)+z(t)}{3}$ (slow variable, convergent strongly) and $t \mapsto (y(t) - x(t), z(t) - y(t))$ (fast variables, convergent in measure) computed by symplectic Euler and the induced symplectic FLAVOR (2.45)). Define $q := (x, y, z)$. To illustrate the F-convergence property of FLAVOR, we fix $H = 1$, vary the mesostep $\delta = H/M$ by changing M, and show in Euclidean norm the difference between $\frac{1}{M}\sum_{i=0}^{M-1} q(T - ih/M)$ computed by FLAVOR and symplectic Euler in Figure 2.5(a). Notice that we average over a short time span of width h, so that the convergence on both the slow variable and the fast variable (convergent only in the sense of measure) could be captured. As shown in Figure 2.5(a), the error scales linearly with $\frac{1}{M}$ for sufficiently large M's, and therefore the global



(a) ***Asymptotically linear error dependence on*** $\delta = 1/M$

(b) ***Asymptotically linear error dependence on total simulation time*** T

(c) ***Asymptotically independent of the scaling factor*** ω

Figure 2.5: Error dependence on parameters in a FLAVOR simulation of (2.100)

error is a linear function of the mesostep δ and the method is first-order uniformly convergent. Figure 2.5(b) shows that the error in general grows linearly with the total simulation time, and this linear growth is observed over a very long simulation time span (even longer than $\omega = \epsilon^{-1/2}$). Figure 2.5(c) shows that the error does not depend on ω ($\epsilon^{-1/2}$) for a fixed δ, as long as ϵ is not too large (i.e., ω not too small). This is not caused by reaching the limit of machine accuracy, but a characteristic of the method: the plateau for large ω corresponds to a almost complete scale separation, where the error of FLAVOR will only depend on the mesostep, which is again intuitive because FLAVOR as a multiscale method is uniformly convergent (despite of ω).

Again, there is no resonant value of δ in the terms of a blown up error.

The fact that the error scales linearly with total simulation time is a much stronger (numerical) result than our (theoretical) error analysis for FLAVORs (in which the error is bounded by a term growing exponentially with the total simulation time). We conjecture that the linear growth of the error is a consequence of the fact that FLAVOR is symplectic and is only true for a subclass of systems, possibly integrable systems. A rigorous analysis of the effects of the structure preservation of FLAVORs on long-term behavior remains to be done.

2.6 Numerical experiments

2.6.1 Hidden Van der Pol oscillator (ODE)

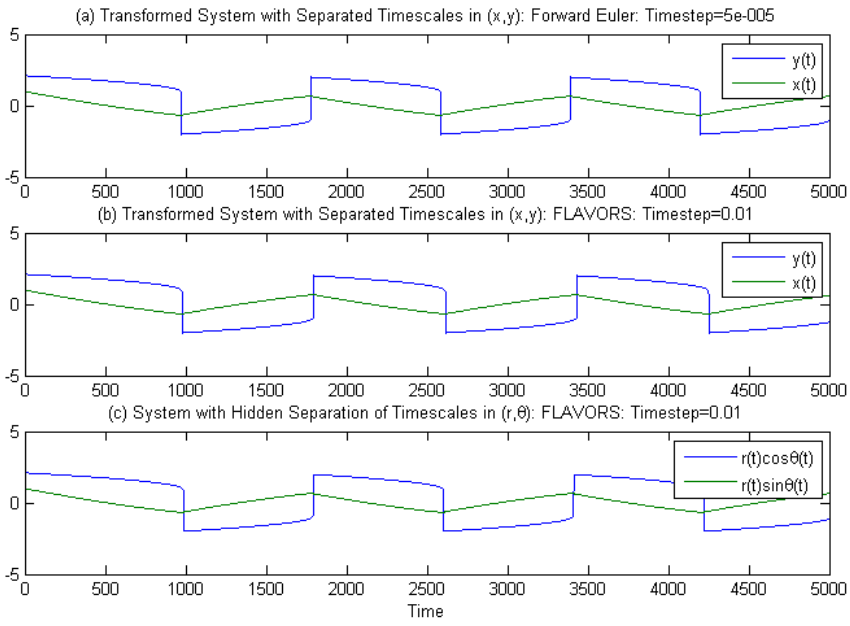


Figure 2.6: Van der Pol oscillator over a timespan of $5/\epsilon$ (a) Direct Forward Euler simulation of (2.102) with time steps resolving the fast time scale (b) (nonintrusive (2.34)) FLAVOR simulation of (2.102) (c) Polar to cartesian image of the (nonintrusive (2.34)) FLAVOR simulation of (2.101) with hidden slow and fast variables. Forward Euler uses time step $h = 0.05\epsilon = 0.00005$. The two FLAVORS simulations use $\delta = 0.01$ and $\tau = 0.00005$. Parameters are $\frac{1}{\epsilon} = 1000$, $x(0) = 1$, $y(0) = 1$

Consider the following system ODEs

$$\begin{cases} \dot{r} = \frac{1}{\epsilon}(r\cos\theta + r\sin\theta - \frac{1}{3}r^3\cos^3\theta)\cos\theta - \epsilon\, r\cos\theta\sin\theta \\ \dot{\theta} = -\epsilon\, cos^2\theta - \frac{1}{\epsilon}(\cos\theta + \sin\theta - \frac{1}{3}r^2\cos^3\theta)\sin\theta \end{cases} \tag{2.101}$$

where $\epsilon \ll 1$. Taking the transformation from polar coordinates to Cartesian coordinates by $[x, y] = [r\sin\theta, r\cos\theta]$ as the local diffeomorphism, we obtained

the hidden system:

$$\begin{cases} \dot{x} = -\epsilon y \\ \dot{y} = \frac{1}{\epsilon}(x + y - \frac{1}{3}y^3) \end{cases} \tag{2.102}$$

Taking the second time derivative of y, the system can also be written as the 2^{nd}-order ODE:

$$\ddot{y} + y = \frac{1}{\epsilon}(1 - y^2)\dot{y}. \tag{2.103}$$

The latter is the classical Van der Pol oscillator [291]. Nonintrusive FLAVOR as defined by (2.34) can be directly applied to (2.101) (with hidden slow and fast processes) by turning on and off the stiff parameter $\frac{1}{\epsilon}$. More precisely, defining $\Phi^{\epsilon,\alpha}(r,\theta)$ by

$$\Phi_h^{\alpha,\epsilon}(r,\theta) := \begin{pmatrix} r \\ \theta \end{pmatrix} + \alpha h \begin{pmatrix} (r\cos\theta + r\sin\theta - \frac{1}{3}r^3\cos^3\theta)\cos\theta \\ -(\cos\theta + \sin\theta - \frac{1}{3}r^2\cos^3\theta)\sin\theta \end{pmatrix} - \epsilon h \begin{pmatrix} r\cos\theta\sin\theta \\ cos^2\theta \end{pmatrix} \tag{2.104}$$

FLAVOR is defined by (2.34) with $\bar{u} := (\bar{r}, \bar{\theta})$, i.e.,

$$(\bar{r}_t, \bar{\theta}_t) = \left(\Phi_{\delta-\tau}^{0,\epsilon} \circ \Phi_{\tau}^{\frac{1}{\epsilon},\epsilon}\right)^k (r_0, \theta_0) \quad \text{for} \quad k\delta \le t < (k+1)\delta. \tag{2.105}$$

We refer to Figure 2.6 for a comparison of integrations by Forward Euler (benchmark) and FLAVORs. FLAVORs give trajectories close to Forward Euler and correctly capture the $\mathcal{O}(\frac{1}{\epsilon})$ period [291] of the relaxation oscillation. Moreover, a 200x acceleration is achieved using FLAVOR.

2.6.2 Hamiltonian system with nonlinear stiff and soft potentials

In this section, we will apply the Symplectic Euler FLAVOR defined by (2.45) and (2.49) to a mechanical system whose Hamiltonian is

$$H(y, x, p_y, p_x) := \frac{1}{2}p_y^2 + \frac{1}{2}p_x^2 + \epsilon^{-1}y^6 + (x - y)^4 \tag{2.106}$$

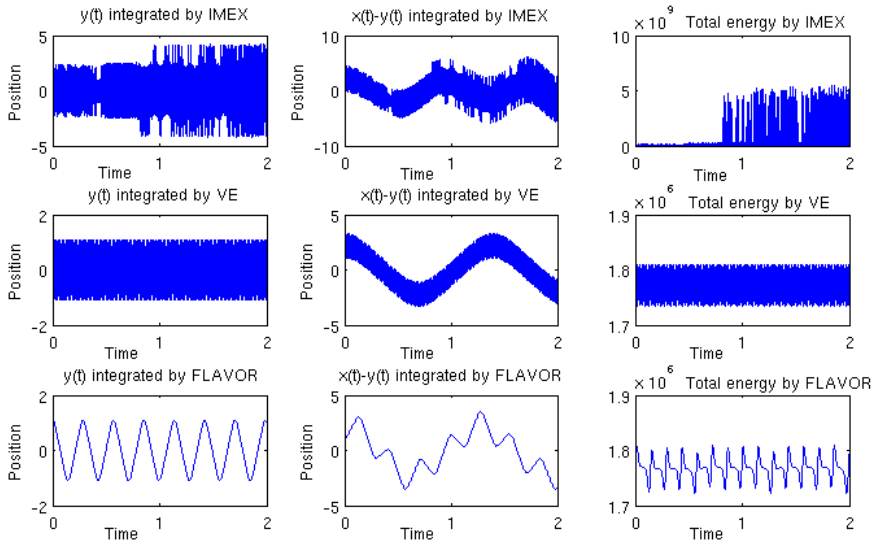


Figure 2.7: In this experiment, $\epsilon = 10^{-6}$, $y(0) = 1.1$, $x(0) = 2.2$, $p_y(0) = 0$ and $p_x(0) = 0$. Simulation time $T = 2$. FLAVOR (defined by (2.45) and (2.49)) uses mesostep $\delta = 10^{-3}$ and microstep $\tau = 10^{-5}$, Variational Euler uses small time step $\tau = 10^{-5}$, and IMEX uses mesostep $\delta = 10^{-3}$. Since the fast potential is nonlinear, IMEX is an implicit method and nonlinear equations have to be solved at every step, and IMEX turns out to be slower than Variational Euler. FLAVOR is strongly accurate with respect to slow variables and accurate in the sense of measures with respect to fast variables. Comparing to Symplectic Euler, FLAVOR accelerated the computation by roughly 100x.

Here, stiff potential $\epsilon^{-1}U = \epsilon^{-1}y^6$ and soft potential $V = (x-y)^4$ are both nonlinear.

Figure 2.7 illustrates $t \mapsto y(t)$ (dominated by a fast process), $t \mapsto x(t) - y(t)$ (a slow process modulated by a fast process), and $t \mapsto H(t)$, respectively computed by Symplectic Euler, the induced symplectic FLAVOR ((2.45) and (2.49)), and IMEX [265]. Notice that $x - y$ is not a purely slow variable but contains some fast component, and therefore the FLAVOR integration of it contains a modulation of local oscillations, which could be interpreted as that fast component slowed down by FLAVOR. It is not easy to find a purely slow variable or a purely fast variable in the form of (2.2) for this example, but the integrated trajectory for such a slow variable will not contain these slowed-down local oscillations.

2.6.3 Fermi-Pasta-Ulam problem over four timescales

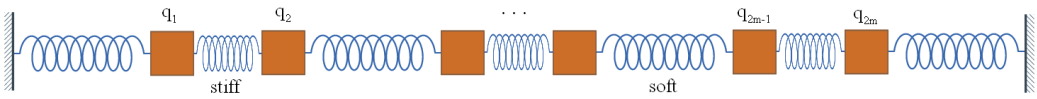
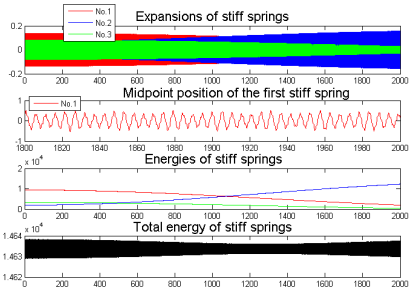


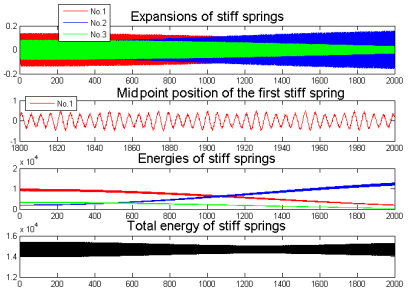
Figure 2.8: Fermi-Pasta-Ulam problem [101] – 1D chain of alternatively connected harmonic stiff and non-harmonic soft springs

In this section, we will consider the Fermi-Pasta-Ulam (FPU) problem [101] illustrated by Figure 2.8, which is mechanical system with the Hamiltonian

$$H(q,p) := \frac{1}{2}\sum_{i=1}^{m}(p_{2i-1}^2 + p_{2i}^2) + \frac{\omega^2}{4}\sum_{i=1}^{m}(q_{2i} - q_{2i-1})^2 + \sum_{i=0}^{m}(q_{2i+1} - q_{2i})^4. \quad (2.107)$$



(a) *By Variational Euler with small time step* $\tau' = 5 \times 10^{-5} = 0.05/\omega$***. 38 periods in Subplot 2 with zoomed-in time axis (∼380 in total over the whole simulation span).***



(b) *By artificial FLAVOR* (2.51) ***with mesostep*** $\delta = 0.002$ ***and microstep*** $\tau = 10^{-4} = 0.1/\omega$***. 38 periods in Subplot 2 with zoomed-in time axis (∼380 in total over the whole simulation span).***

Figure 2.9: Simulations of the FPU problem over $T = 2\omega$. Subplot 2 of both figures have zoomed-in time axes so that whether phase lag or any other distortion of trajectory exists could be closely investigated. In this experiment $m = 3$, $\omega = 10^3$, $x(0) = [0.4642, -0.4202, 0.0344, 0.1371, 0.0626, 0.0810]$ is randomly chosen, and $y(0) = [0, 0, 0, 0, 0, 0]$.

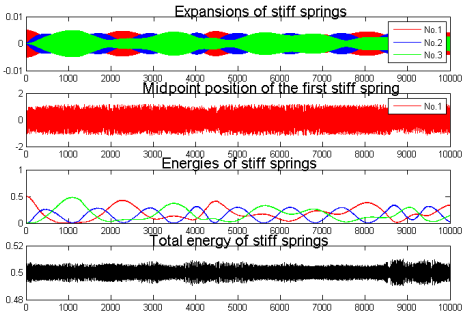
The FPU problem is a well known benchmark problem [201, 128] for multiscale integrators because it exhibits different behaviors over widely separated timescales. The stiff springs nearly behave like harmonic oscillators with period $\sim \mathcal{O}(\omega^{-1})$. Then, the centers of masses linked by stiff springs (i.e., the midpoints of stiff

springs) change over a timescale $\mathcal{O}(1)$. The third timescale, $\mathcal{O}(\omega)$, is associated with the rate of energy exchange between stiff springs. Energy exchange among stiff springs extends to even slower timescales, in either a periodic or a chaotic fashion [106, 104]. On the other hand, the total energy of the stiff springs behaves almost like a constant over an even longer time span. This wide separation of timescales can be seen in Figures 2.9, 2.10, and 2.12, where four subplots address different scales: Subplot 1 shows the fast variables $(q_{2i} - q_{2i-1})/\sqrt{2}$; Subplot 2 shows one of the slow variables $(q_2 + q_1)/\sqrt{2}$; Subplot 3 shows the energy transfer pattern among stiff springs, which is even slower; Subplot 4 shows the near-constant total energy of three stiff springs. All four subplots are time-series. Comprehensive surveys on FPU problem, including discussions on timescales and numerical recipes, can be found in [128, 71].

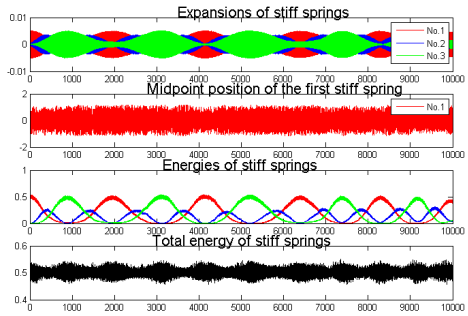
Figures 2.9(a) and 2.9(b) compare symplectic Euler (with microsopic time steps) with the artificial FLAVOR (2.51). On a timescale $\mathcal{O}(\omega)$ ($\omega \gg 1$), FLAVOR captured slow variable's periodic behavior with the correct period and phase, as well as the slower process of energy transfer. At the same time, FLAVOR accelerated the computation by roughly 40x (since $\delta = 40\tau'$).

It is not worrisome that artificial FLAVOR produces stiff spring energy trajectories with rapid local oscillations, which exhibit both thicker individual energy curves and total energy with larger variance. In fact, these local oscillations do not seem to affect the global transfer pattern nor its period and are caused by the numerical error associated with microstep τ. This can be inferred because local oscillations disappear after replacing the Variational Euler approximation of θ_τ^ϵ by the exact flow of H^{fast}. As illustrated in Figure 2.11, the exact flow helps to obtain thin energy curves of stiff springs with no rapid local oscillations, as well as a total energy with a variance even smaller than that given by fine-step Variational Euler (Figure 2.9(a)).

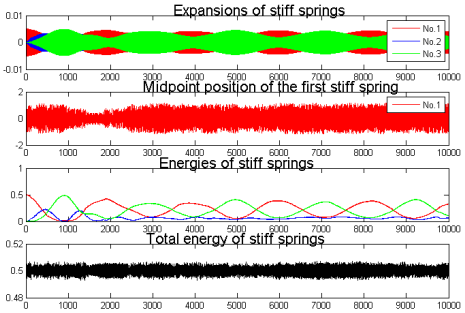
Now, we reach further to $\mathcal{O}(\omega^2)$ total integration time to investigate different integrators' performances in capturing the long time energy exchange pattern (Figure 2.10).



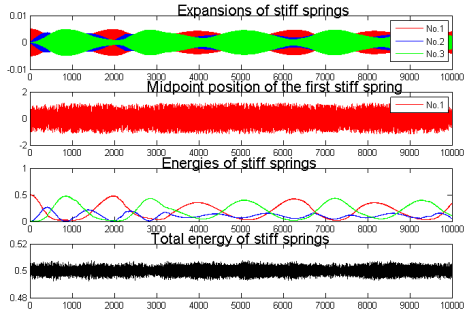
(a) ***By Velocity Verlet with tiny time step*** $h = 10^{-5}$ ***(two orders of magnitude smaller than the stability limit).***



(b) ***By artificial FLAVOR*** (2.51) ***with mesostep*** $\delta = 0.002$ ***and microstep*** $\tau = 0.0005 = 0.1/\omega$**.**



(c) ***By IMEX with mesostep*** $\delta = 0.002$**.**



(d) ***By Impulse Method with mesostep*** $\delta = 0.002$**.**

Figure 2.10: Simulations of FPU problem over $T = \frac{1}{4}\omega^2$. Initial conditions are $x(0) = [1, 0, 0, 1/\omega, 0, 0]$ and $y(0) = [0, 0, 0, 0, 0, 0]$ so that energy starts concentrated on the leftmost soft and stiff springs. We chose a smaller $\omega = 200$ because with a larger ω it would take weeks to run Velocity Verlet on a laptop.

There is a significant difference among stiff spring energy transfer patterns produced by Velocity Verlet, FLAVOR, IMEX and the Impulse Method. Here, there is no analytic solution or provably accurate method for comparison. FLAVOR is the only method that shows periodic behavior on the long time scale and convergence tests show that FLAVOR's trajectories remain stable under small variations of step sizes.

Notice that the system would be integrable and periodic if nonlinearity did not exist (see Figure 2.12 for integration of a system in which the slow potential

is quadratic; there, the system could be perfectly integrated by all FLAVORs, Velocity-Verlet, IMEX and the Impulse Method (results not shown)). When nonlinearity is present, for a fixed small number of springs, the nonlinearity will not destroy the periodicity of the system unless the linear counterpart is weak (i.e., ω is small; see for instance [186] for an example of chaotic threshold on nonlinearity), which should not be the case here since ω is very large.

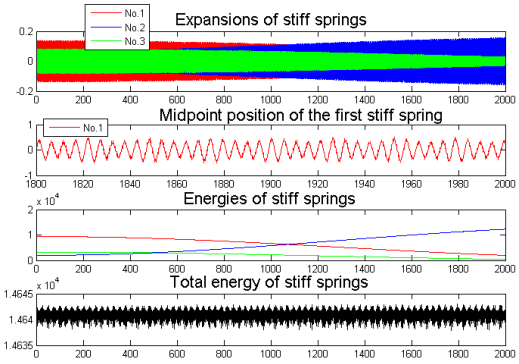


Figure 2.11: By artificial FLAVORs (Section 2.2.6) based on exact fast flow with mesostep $\delta = 0.002$ and microstep $\tau = 10^{-4}$. Less oscillatory stiff spring energies. 38 periods in Subplot 2 with zoomed-in time axis (∼380 in total over the whole simulation span).

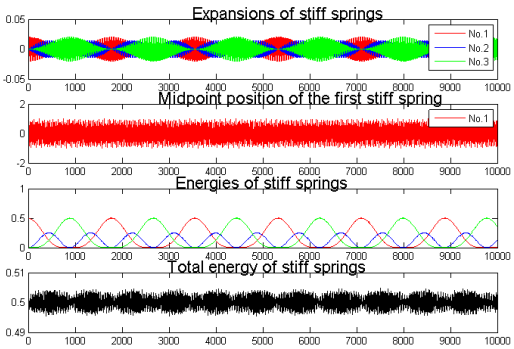
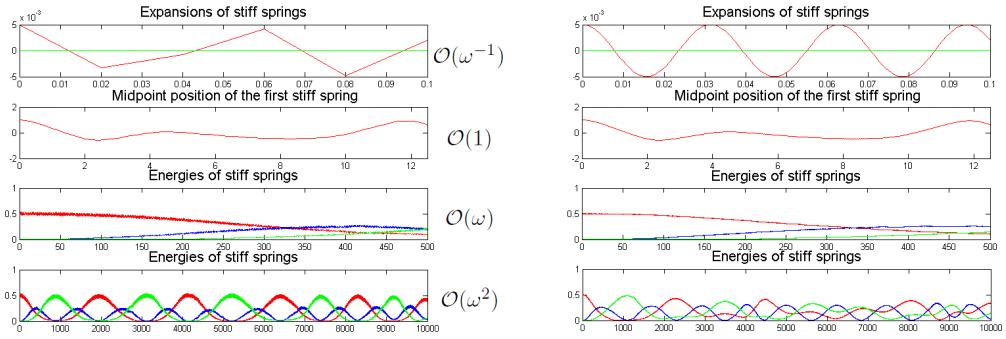


Figure 2.12: Harmonic FPU, $T = 50\omega$, exact solution

Since we employed a tiny timestep in Velocity-Verlet, originally in hope that it could be used as a long time simulation benchmark, it is worth discussing why its performance is still not satisfactory. Being a second-order method, Velocity-Verlet has an error bound of $\mathcal{O}(e^T h^2)$. On the other hand, backward error analysis guarantees that the energy of the integrated trajectory oscillates around the true conserved energy, hence eliminating the possibility of exponential growth of the numerical solution. Nevertheless, at this moment there is no result known to the authors to link these two analytical results to guarantee long term accuracy on the exchange of stiff springs' energies. This exchange is in fact a delicate phenomenon, and a slight distortion in stiff spring lengths could easily disrupt its period or even its periodicity. We believe this is what Velocity-Verlet numerical errors did in the

long time simulation.

These numerical observations seem to indicate that symplectic FLAVORs may have special long time properties. Specifically, although we could not quantify the error here because there is no benchmark to compare to when the total simulation time is $\mathcal{O}(\omega^2)$, the long term behavior seems to indicate an error growing much slower than exponentially (please refer to Remark 2.1.7 for a discussion on exponential error bounds and Figure 2.5(b) for another example of conjectured linear error growth). A rigorous investigation on FLAVORs' long time behavior remains to be done.



(a) *FLAVOR* **(b) *Velocity Verlet***

Figure 2.13: Quantities of interest in integrations of FPU over different timescales. FLAVOR (2.51) captures the fastest timescale in the sense of measure, while Velocity Verlet cannot accurately capture the slowest ($\mathcal{O}(\omega^2)$) timescale despite the small time step it uses. Here FLAVOR is 200 times faster than Velocity Verlet. All parameters are the same as in Figure 2.10(a) and 2.10(b), e.g., $\omega = 200$, $\delta = 0.002$, $\tau = 0.0005$ and $h = 10^{-5}$.

Figure 2.13 summarizes FLAVOR's performance on various timescales in a comparison to Velocity Verlet.

Notice that there are many sophisticated methods designed to integrate the FPU problem (see [128] for a review), as well as general multiscale methods that can be applied to the FPU problem. HMM uses an identification of slow variables [13] to capture the energy transfer between stiff springs over a time span at the order of ω. Simulations shown here, however, are over a much longer time span at the order of $\mathcal{O}(\omega^2)$.

Remark 2.6.1 (On resonances)**.** *Multiscale in time integrators are usually plagued by two kinds of resonances. FLAVORs have none of them.*

The first type, called Takens resonance [271], is related to the case in which there are no closed equations for only slow variables [38]. FLAVORs avoid Takens resonance because they do necessarily not look for the closed equations, but instead keep the information on the local invariant measure of fast variables (recall the notion of F-convergence in Section 2.1.3). Observe that the FPU problem exhibits Takens resonance because the eigenfrequencies of the strong potential are identical. Nevertheless, FLAVORs still capture the solution trajectories given any large value of ω *with mesostep* $\delta \gg 1/\omega$ *independent of* ω*.*

The second type [55] is related to instabilities created by interactions between parameters ϵ*,* τ*,* δ*, and the intrinsic frequency of the system. For instance, if* $\epsilon^{-1} = \omega^2$*, one suspects that resonances could happen at* $\omega\delta$ *or* $\omega\tau$ *equal to multiples of* $\pi/2$*. The analysis provided in Section 2.5 shows that such an unstable interaction does not occur, either in the sense of stability or in terms of numerical errors. This can be intuitively understood upon observing that FLAVORs never approximate* $\cos(\delta\omega)$*, while on the other hand, they do approximate* $\cos(\tau\omega)$*, whose resonance frequency* $\tau = 2\pi/\omega$ *is ruled out by the requirement that* $\tau \ll \epsilon$ *for nonintrusive FLAVOR and* $\tau \ll \sqrt{\epsilon}$ *for artificial FLAVOR.*

2.6.4 Nonlinear two-dimensional primitive molecular dynamics

Now consider a two-dimensional, two degrees of freedom example in which a point mass is linked through a spring to a massless fixed hinge at the origin. While the spring as well as the point mass are allowed to rotate around the hinge (the spring remains straight), the more the spring-mass tilts away from its equilibrium angle the more restorative force it will experience. This example is a simplified version of prevailing molecular dynamics models, in which bond lengths and angles between neighboring bonds are both spring-like; other (non-local) potential energy terms are ignored.

Denote by x and y the Euclidean coordinates of the mass, and p_x, p_y the

corresponding momenta. Also, introduce polar coordinates (r, θ), with $x = r\cos\theta$ and $y = r\sin\theta$. Then the Hamiltonian reads

$$\begin{aligned} H &= \frac{1}{2}p_x^2 + \frac{1}{2}p_y^2 + \frac{1}{2}\omega^2(r - r_0)^2 + (\cos\theta)^2 \\ &= \frac{1}{2}p_x^2 + \frac{1}{2}p_y^2 + \frac{1}{2}\omega^2(\sqrt{x^2+y^2} - r_0)^2 + \frac{x^2}{x^2+y^2} \end{aligned} \tag{2.108}$$

where r_0 is equilibrium bond length parameter, and ω is a large number that indicates the bond oscillation frequency.

Remark 2.6.2. *This seemingly trivial example is not easy to integrate.*

1. *If the system is viewed in Euclidean coordinates (x, y, p_x, p_y) it is completely nonlinear with a nonpolynomial potential, and hence the Impulse Method or its variations [124, 286, 109, 240], or IMEX [265], or the homogenization method introduced in [176] cannot be applied using a mesostep.*

2. *If the Hamiltonian is rewritten in generalized coordinates $(r, \theta, p_r, p_\theta)$, $H = \frac{1}{2}p_r^2 + \frac{1}{2}\frac{p_\theta^2}{r^2} + \frac{1}{2}\omega^2(r-r_0)^2 + \frac{1}{2}\cos(\theta)^2$, a fast quadratic potential can be identified. However, the mass matrix $\begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}$ is not constant, but rapidly oscillating, and hence methods that work for quasi-quadratic fast potentials (i.e., "harmonic oscillator" with a slowly changing frequency) ([176] for example) cannot be applied.*

Figure 2.14 compares symplectic Euler with the induced symplectic FLAVOR ((2.45) and (2.49)) applied to (2.108) in Euclidean coordinates. Also, the implementation of an artificial FLAVOR on this example (if needed) is easy, because the free dynamics with no soft potential and frozen stiff potential is just a rotation around the origin.

FLAVOR reproduced the slow θ trajectory and accelerated the simulation by roughly 50x times (since $\delta = 50\tau$). It can also be seen from both energy fluctuations

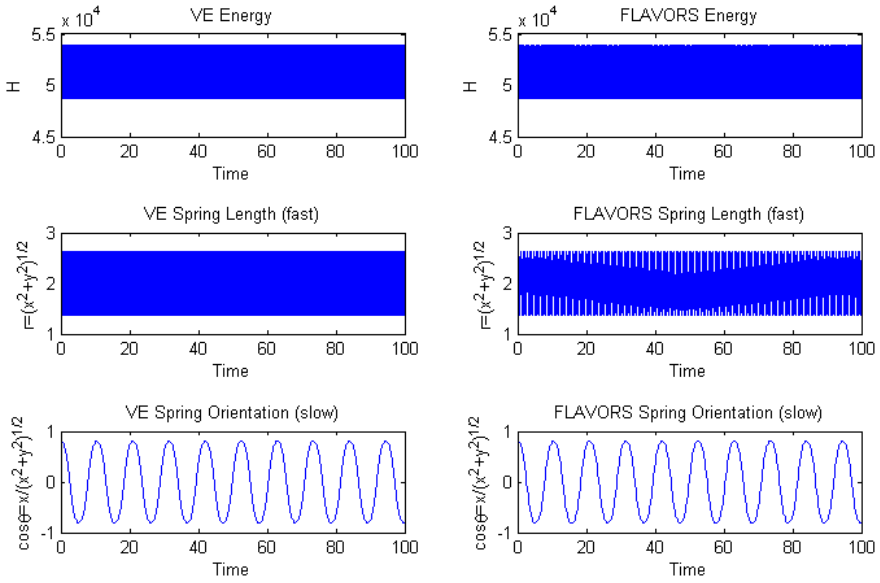


Figure 2.14: Simulation of (2.108). Symplectic Euler uses small time step $\tau = 0.0002$ and the induced symplectic FLAVOR ((2.45) and (2.49)) uses mesostep $\delta = 0.01$ and microstep $\tau = 0.0002$. In this simulation $\omega = 500$, $x(0) = 1.1$, $y(0) = 0.8$, $p_x(0) = 0$, $p_y(0) = 0$ and simulation time $T = 100$.

and the trajectory of the fast variable that the fast process' amplitude is well captured although its period has been lengthened.

2.6.5 Nonlinear molecular clip

We now consider a united-atom representation of a three-atom polymer with two bonds (e.g., propane or water molecule). This is a simplified version of several prevailing molecular dynamics force fields (for example, CHARMM [51], AMBER [76], or a simpler example of butane [235, 238]). Since the angular momentum is conserved, instead of the 3D space we could fix the coordinate system in a 2D plane. Introduce both Cartesian coordinates $(x_1, y_1, x_2, y_2, x_3, y_3)$ and generalized coordinates $r_1 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ and $r_2 = \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}$ for bond lengths and θ for the angle between the two bonds (Figure 2.15). The kinetic

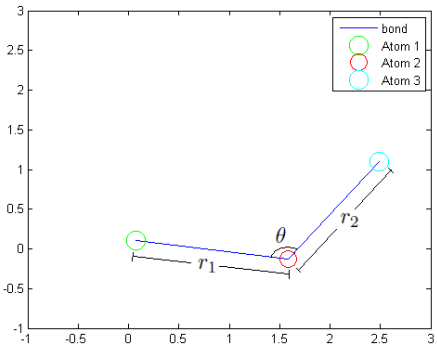


Figure 2.15: One example configuration of a propane molecule (united-atom representation).

energy is

$$K.E. = \frac{1}{2}m_1(\dot{x}_1^2 + \dot{y}_1^2) + \frac{1}{2}m_2(\dot{x}_2^2 + \dot{y}_2^2) + \frac{1}{2}m_3(\dot{x}_3^2 + \dot{y}_3^2) \tag{2.109}$$

where m_1, m_2, and m_3 denote the masses of the atoms.

The potential energy consists of a bond term and a bond angle term, both of which are of harmonic oscillator type:

$$P.E. = V_{bond} + V_{angle} \tag{2.110}$$

$$V_{bond} = \frac{1}{2}K_r[(r_1 - r_0)^2 + (r_2 - r_0)^2] \tag{2.111}$$

$$V_{angle} = \frac{1}{2}K_\theta(\cos(\theta) - \cos(\theta_0))^2 \tag{2.112}$$

Notice that the system is in fact fully nonlinear: if written in generalized coordinates, the kinetic energy will correspond to a nonlinear and position-dependent mass matrix, whereas in Cartesian coordinates, both terms in the potential energy are non-polynomial functions of the configuration.

In the case of propane, $m_1 = 15\mu, m_2 = 14\mu, m_3 = 15\mu$ where $\mu = 1.67 \cdot 10^{-27}kg$, $r_0 = 1.53\mathring{A}$, $K_r = 83.7kcal/(mol\mathring{A}^2)$, $\theta_0 = 109.5°$ and $K_\theta = 43.1kcal/mol$ [235].

The propane system is characterized by a separation of timescales to some

extent: bond stretching and bond-angle bending are characterized by 10^{14} and 10^{13} Hz vibrational frequencies respectively [304]. To examine FLAVORs, we use unitless parameters and exaggerate the timescale separation by setting K_r to be 8370 and K_θ to be 4.31. We also let $\mu = 1$, without loss of generality, for arithmetic considerations.

In this system, the bond potential is the fast potential and the bond-angle potential is the slow one. It is well known that using only a coarse time step (characteristic of bond-angle oscillations) by freezing bond lengths produces biased results, and many physics-based methods have been proposed to remedy this difficulty (for example the approach of Fixman [103]; also see a review in [304]). On the other hand, few multiscale methods work for this fully nonlinear system.

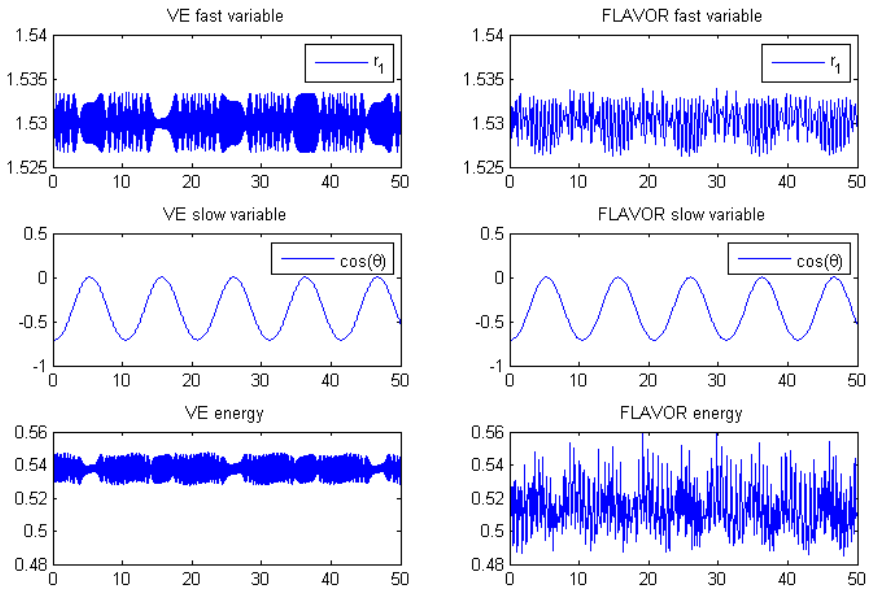


Figure 2.16: Simulations of exaggerated propane molecule (Section 2.6.5). Symplectic Euler uses $h = 0.01$ and the induced symplectic FLAVOR ((2.45) and (2.49)) parameters are $\delta = 0.1$ and $\tau = 0.01$. Initial conditions are $[x_1, y_1, x_2, y_2, x_3, y_3] = [0, 0, 1.533, 0, 2.6136, 1.0826]$ and $[m_1\dot{x}_1, m_1\dot{y}_1, m_2\dot{x}_2, m_2\dot{y}_2, m_3\dot{x}_3, m_3\dot{y}_3] = [-0.4326, -1.6656, 0.1253, 0.2877, -1.1465, 1.1909]$ (randomly chosen).

Figure 2.16 compares symplectic Euler with the induced symplectic FLA-

VOR ((2.45) and (2.49)) applied in Euclidean coordinates. 10x acceleration is achieved. A simulation movie is also available at `http://www.youtube.com/watch?v=BxMIdE_FN1k` .

2.6.6 Forced nonautonomous mechanical system: Kapitza's inverted pendulum

As the famous Kapitza's inverted pendulum [162] shows (for recent references, see for instance [14] for a numerical integration, and [241] for a generalization to a stochastic setting), the up position of a single pendulum can be stabilized if the pivot of the pendulum experiences external forcing in the form of vertical oscillation. Specifically, if the position of the pivot is given by $y = sin(\omega t)$, the system is governed by

$$l\ddot{\theta} = [g + \omega^2 \sin(2\pi\omega t)] \sin\theta \tag{2.113}$$

where θ denotes the clockwise angle of the pendulum from the positive y direction, l is the length of the pendulum and g is the gravitational constant. In this case, the rapid vibration causes the pendulum to oscillate slowly around the positive y direction with a $\mathcal{O}(1)$ frequency.

A single scale integration of this system could be done by Variational Euler with discrete d'Alembert principle to account for external forces [192]:

$$\begin{cases} \theta_{(i+1)h} = \theta_{ih} + hp_{ih}/l \\ p_{(i+1)h} = p_{ih} + h[g + \omega^2 \sin(2\pi\omega ih)] \sin(\theta_{(i+1)h}) \end{cases}, \tag{2.114}$$

where the time step h has to be smaller than $\mathcal{O}(1/\omega)$.

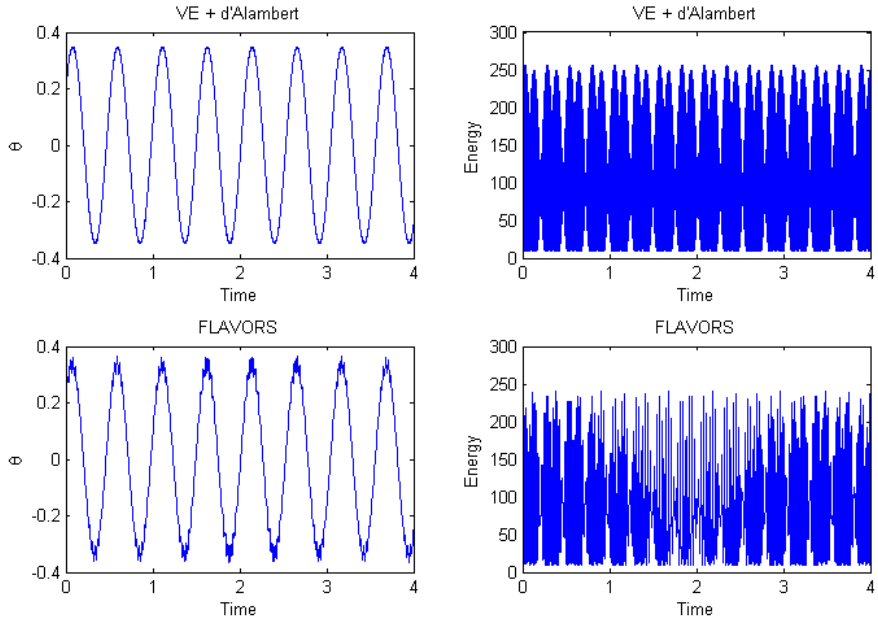


Figure 2.17: Simulations of the inverted pendulum. The integration by Variational Euler + d'Alembert principle uses time step $h = 0.2/\omega/\sqrt{l} \approx 0.000067$, while FLAVOR (defined by (2.115)) uses $\delta = 0.002$ and $\tau = 0.2/\omega/\sqrt{l}$. Also, $g = 9.8$, $l = 9$, $\theta(0) = 0.2$, $\dot{\theta}(0) = 0$ and $\omega = 1000$

A FLAVOR is given by

$$\begin{cases} \theta_{n\delta+\tau} = \theta_{n\delta} + \tau p_{n\delta}/l \\ p_{n\delta+\tau} = p_{n\delta} + \tau[g + \omega^2 \sin(2\pi\omega n\tau)] \sin(\theta_{n\delta+\tau}) \\ \theta_{(n+1)\delta} = \theta_{n\delta+\tau} + (\delta - \tau) p_{n\delta+\tau}/l \\ p_{(n+1)\delta} = p_{n\delta+\tau} + (\delta - \tau) g \sin(\theta_{(n+1)\delta}) \end{cases} \tag{2.115}$$

Observe that the time dependent force is resynchronized to the τ time scale from the δ time scale. Specifically, the FLAVOR (2.115) uses $\omega^2 \sin(2\pi\omega n\tau)$ instead of $\omega^2 \sin(2\pi\omega n\delta)$.

Numerical results are illustrated in Figure 2.17 (also available as a movie at `http://www.youtube.com/watch?v=QL2oFq9fyXM`). Notice in this example that

θ, being the only degree of freedom, contains a combination of slow and fast dynamics. FLAVOR could only capture the fast dynamics in the sense of measures, and this is why dents appear as modulation on the slow oscillation of θ. Also, although this forced system does not admit a conserved energy, the value of the Hamiltonian should oscillate periodically due to the periodic external driving force. A non-mechanics based method (e.g., Forward Euler) would produce an unbounded growth or a decrease in the energy, which is wrong, but FLAVORs do not have this drawback.

2.6.7 Nonautonomous SDE system with hidden slow variables

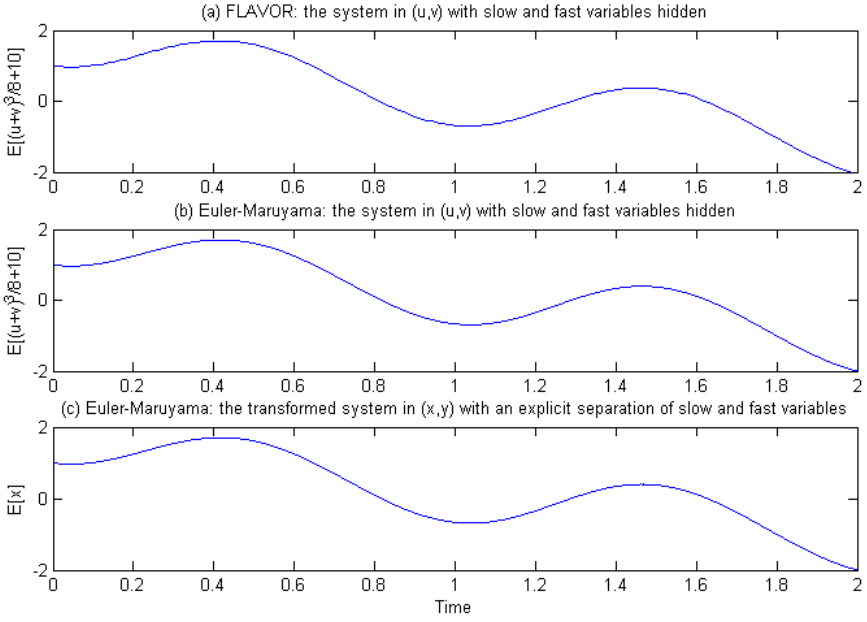


Figure 2.18: (a) Integration of (2.116) by nonintrusive FLAVOR (2.65) using mesostep step $\delta = 0.01$ (b) Integration of (2.116) by Euler-Maruyama using fine time step $h = 10^{-4}$ (c) Integration of (2.118) by Euler-Maruyama using the same small step $h = 10^{-4}$. Expectations of the slow variable (whether or not hidden) are obtained by empirically averaging over an ensemble of 100 independent sample trajectories. $\epsilon = 10^{-4}$, $x(0) = 1 + \epsilon$, $y(0) = 1$, $T = 2$ (the expectation of the real solution will blow up around $T = 3$). We have chosen a big enough $c = 10$ so that the scale separation transformation is a diffeomorphism.

Consider the following artificially made nonautonomous SDE system

$$\begin{cases} du = \frac{4}{3(u+v)^2}\left(-\frac{1}{2}\left(\frac{v-u}{2}\right)^2 + 5\sin(2\pi t)\right)dt - \frac{1}{\epsilon}\left(\left(\frac{u+v}{2}\right)^3 + c - \frac{v-u}{2}\right)dt - \sqrt{\frac{2}{\epsilon}}dW_t \\ dv = \frac{4}{3(u+v)^2}\left(-\frac{1}{2}\left(\frac{v-u}{2}\right)^2 + 5\sin(2\pi t)\right)dt + \frac{1}{\epsilon}\left(\left(\frac{u+v}{2}\right)^3 + c - \frac{v-u}{2}\right)dt + \sqrt{\frac{2}{\epsilon}}dW_t \end{cases} \tag{2.116}$$

where c is a constant and the two dW_t terms refer to the same Brownian motion. The system (2.116) can be converted via the local diffeomorphism

$$\begin{cases} u = (x-c)^{1/3} - y \\ v = (x-c)^{1/3} + y \end{cases}, \tag{2.117}$$

into the following hidden system separating slow and fast variables

$$\begin{cases} dx = -\frac{1}{2}y^2 dt + 5\sin(2\pi t)dt \\ dy = \frac{1}{\epsilon}(x-y)dt + \sqrt{\frac{2}{\epsilon}}dW_t \end{cases}. \tag{2.118}$$

Nonintrusive FLAVOR (2.65) can be directly applied to (2.116) using a time step $\delta \gg \epsilon$ without prior identification of the slow and fast variables, i.e., without knowing (2.118). The expected values of solutions of (2.116) integrated by FLAVORs with mesostep δ and Euler-Maruyama with a small time step τ are presented in Figure 2.18. FLAVOR has accelerated the computation by 100x.

2.6.8 Langevin equations with slow noise and friction

In this section, we put the double spring system in Section 2.6.2 in the real world (i.e., with noise and friction), and the system is now modeled by the SDEs (both springs are made quartic just for the computational concern on the single-scale

benchmark integration):

$$\begin{cases} dy = p_y dt \\ dx = p_x dt \\ dp_y = -\epsilon^{-1} y^3 dt - 4(y-x)^3 dt - c p_y dt + \sigma dW_t^1 \\ dp_x = -4(x-y)^3 dt - c p_x dt + \sigma dW_t^2 \end{cases} . \tag{2.119}$$

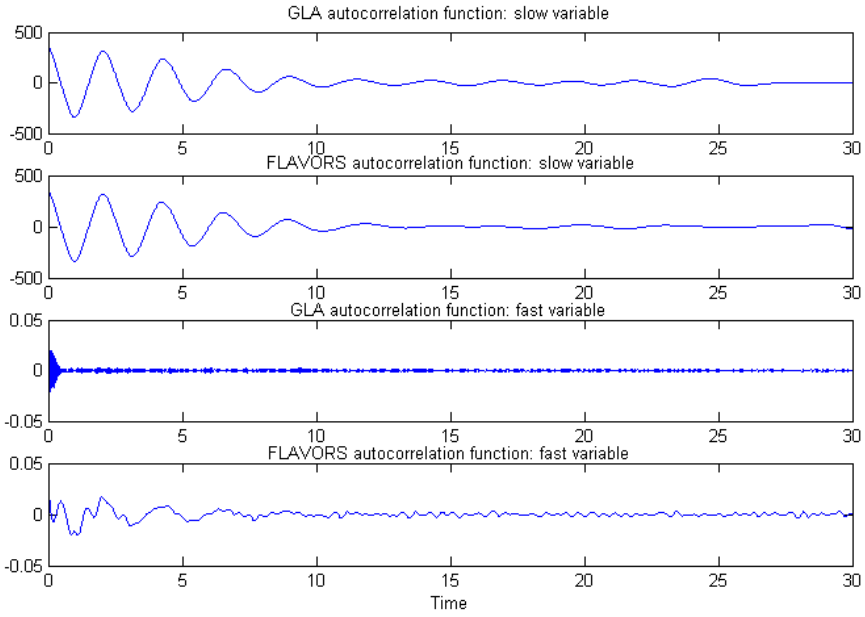
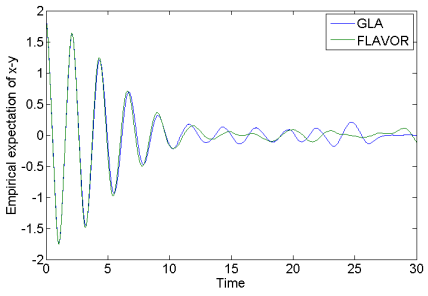
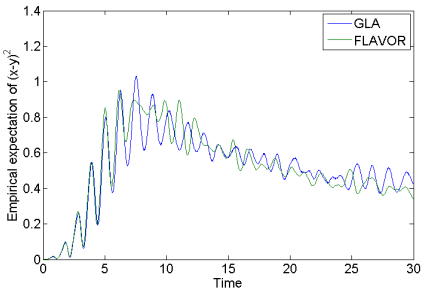


Figure 2.19: SDE (2.119): autocorrelation functions of $\mathbb{E}[y(t)y(0)]$ (dominantly fast) and of $\mathbb{E}[(x(t)-y(t))(x(0)-y(0))]$ (dominantly slow), empirically obtained by GLA and FLAVORs.

We compare several autocorrelation functions and time-dependent moments of this stochastic process integrated by a quasi-symplectic FLAVOR ((2.88) and (2.49)) and Geometric Langevin Algorithm (GLA) [41]. Expectations are empirically calculated by averaging over an ensemble of 100 sample trajectories with $T = 30$, $\epsilon = 10^{-8}$, $\tau = 0.001$, $\delta = 0.01$. $y(0) = 2.1/\omega$ (with $\omega := 1/\sqrt{\epsilon}$), $x(0) = y(0) + 1.8$, $c = 0.1$ and $\sigma = 0.5$. GLA uses time step $h = 0.001$. Noise and



(a) $\mathbb{E}\big(x(t) - y(t)\big)$



(b) $\mathbb{E}\big((x(t) - y(t))^2\big)$

Figure 2.20: SDE (2.119): Empirical moments obtained from simulations of ensembles of (2.119) with GLA and quasi-symplectic FLAVOR (Section 2.4.4)

friction are slow here in the sense that they are not of order $\mathcal{O}(\omega)$ or larger.

As can be seen from Figure 2.19, 2.20(a), and 2.20(b), in the regime dominated by deterministic dynamics (roughly from $t = 0$ to $t = 8$), various moments calculated empirically by FLAVORs and GLA are in agreement. Also in that regime, autocorrelation functions of the slow variables agree, whereas autocorrelation functions of the fast variables agree only in the sense of measures (after time averaging over a mesoscopic ($o(1)$) time span). The discrepancy between FLAVOR and GLA when $t > 8$ is due to stochastic fluctuation and is an effect of the finite number of samples (100) used to compute sample averages.

These numerical results illustrate the statistical properties of FLAVORs in addition to the weak convergence of single trajectory (recall that if the noise is applied to slow variables, FLAVORs do not converge strongly but only in the sense of distributions). Although FLAVORs do not converge to the Boltzmann-Gibbs that GLA converges to (if we just consider the usual sense of total variational norm, the fast variable will have a different marginal distribution), we nevertheless do see that moments of the slow variable converge to the benchmark, which suggests at least a convergence in distribution towards the marginal of the invariant/ergodic distribution that corresponds to the slow variable.

2.6.9 Langevin equations with fast noise and friction

Consider a stochastic mechanical system with a similar configuration to the above. The difference is that the soft spring oscillates at a frequency nonlinearly dependent on the stiff spring's length, and the left mass experiences strong friction and noise while the right mass does not (i.e., degenerate noise). The Hamiltonian is

$$H(y,x,p_y,p_x) = \frac{1}{2}p_y^2 + \frac{1}{2}p_x^2 + \frac{1}{4}\omega^4 y^4 + e^y(x-y)^2, \tag{2.120}$$

and the governing SDEs are:

$$\begin{cases} dy = p_y dt \\ dx = p_x dt \\ dp_y = -\omega^4 y^3 dt - (2+y-x)(y-x)e^y dt - \omega^2 c p_y dt + \omega\sigma dW^t \\ dp_x = -2(x-y)e^y dt \end{cases}. \tag{2.121}$$

In this system, the deterministic dynamics and the effects of noise and friction both involve a $\mathcal{O}(1/\omega^2)$ timescale. We have implemented the stiff noise and friction version of FLAVORs ((2.89) and (2.49)).

In Figure 2.21, we have plotted the first and second moments of the dominantly slow variable $x(t)-y(t)$ as well as the first moment of the dominantly fast variable $y(t)$ as functions of time. Moments of the dominantly slow variable integrated by quasi-symplectic FLAVOR (Section 2.4.4) and GLA [41] concur, numerically suggesting weak convergence and conservation of marginal Boltzmann-Gibbs. 100x computational acceleration is achieved.

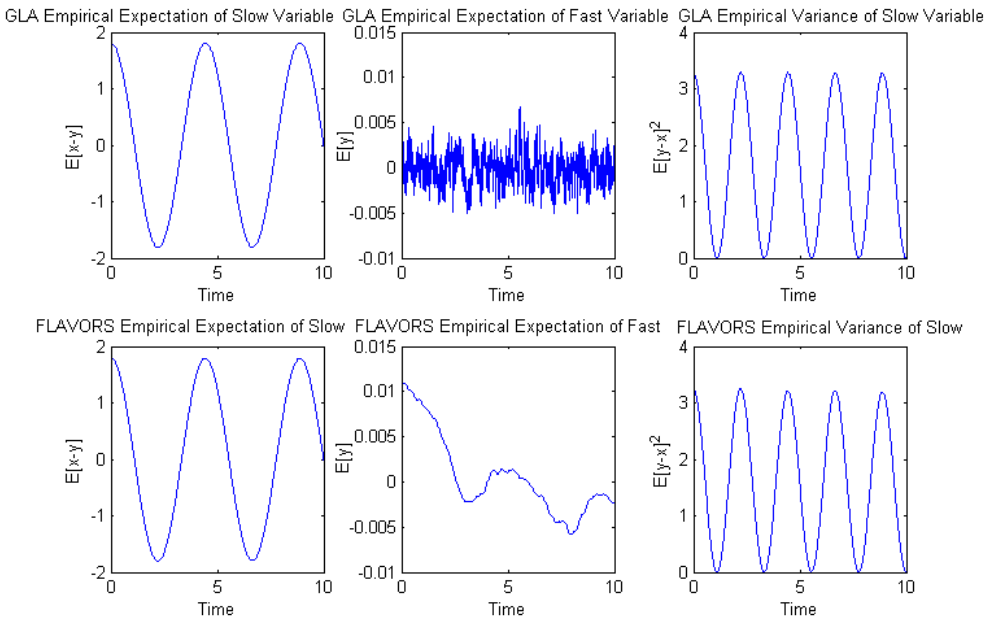


Figure 2.21: $\mathbb{E}[x(t) - y(t)]$, $\mathbb{E}[y(t)]$, and $\mathbb{E}[x(t) - y(t)]^2$ obtained by GLA and quasi-symplectic FLAVOR (Section 2.4.4). Expectations are empirically calculated by averaging over an ensemble of 50 sample trajectories with $T = 10$, $\omega = 100$, $\tau = 10^{-4}$, $\delta = 0.01$. $y(0) = 1.1/\omega$, $x(0) = y(0) + 1.8$, $c = 0.1$ and $\sigma = 1$. GLA uses time step $h = 10^{-4}$.

Chapter 3

FLAVORs for PDEs

We generalize FLow AVeraging integratORs to stiff PDEs (the ODE/SDE version of FLAVORs is introduced in Chapter 2). Again, the strategy of turning on and off stiff coefficients on an alternating microscopic-mesoscopic mesh is adopted. Like in the previous chapter, slow and fast variables do not need to be identified.

The generality of this strategy is illustrated by its applications to finite difference methods (Section 3.1), multi-symplectic integrators (Section 3.2), and pseudospectral methods (Section 3.3); although we have not done so, the proposed strategy can also apply to finite element methods or finite volume methods in order to preserve various structures in multiscale integrations.

The convergence of PDE-FLAVORs is analyzed in Section 3.4.1 via a semi-discrete approach, in which the space is first discretized/interpolated, and then the ODE-FLAVORs convergence result is linked to the error analysis of PDEs. A non-asymptotic error bound is given to quantify the two-scale convergence of the numerical solution (strong convergence on the hidden slow variables and weak convergence on the hidden fast variables).

We also show in Section 3.5 that applying the ODE-FLAVOR strategy to characteristics leads to accurate approximations of stiff PDEs solutions.

Recall (from Section 1.3.2) that the majority of generic multiscale PDE solvers either require identified slow/fast variables (which are in general much more difficult to find because they may depend on the discretization), or rely on assumptions

such as a fast process convergent to a point distribution. FLAVORs do not have these restrictions. Moreover, FLAVORs are among the few approaches to multiscale structure-preserving PDE integrations, because FLAVORs inherit structure preservation properties from legacy codes, such as multisymplecticity illustrated in Section 3.2.

Most of the results in this chapter are published in [278].

3.1 Finite difference and space-time FLAVOR mesh

3.1.1 Single-scale method and limitation

Consider a multiscale PDE:

$$F(1, \epsilon^{-1}, x, t, u(x,t), u_x(x,t), u_t(x,t), u_{xx}(x,t), u_{xt}(x,t), u_{tt}(x,t), \ldots) = 0 \quad (3.1)$$

where F is a given function (possibly nonlinear), ϵ is a small positive real parameter and x and t are spatial and temporal coordinates.

To obtain a numerical solution of (3.1), the simplest single-scale finite difference approach employs a uniform rectangular mesh with time step length h and space step length k, and approximates the solution u by its values at discrete grid points. Differential operators will be approximated by finite differences; for instance, according to forward space forward time rules: $u_x(ik, jh) \approx (u_{i+1,j} - u_{i,j})/k$ and $u_t(ik, jh) \approx (u_{i,j+1} - u_{i,j})/h$, where u_{ij} is the numerical solution at discrete grid point with space index i and time index j. After this discretization, the original PDE is approximated by a finite dimensional algebraic system, which can be solved to yield the numerical solution.

Of course, a necessary condition for obtaining stability and accuracy in the numerical solution is that h and k have to be small enough. A quantitative statement on how small they need to be will depend on the specific PDE and discretization. For 1D linear advection equations $u_x - au_t = 0$ and forward time forward space discretizations, the $h < k/a$ CFL condition [77] has to be met to ensure stabil-

ity, which is also a neccessary condition for accuracy [174]. Intuitively, the CFL condition guarantees that information does not propagate faster than what the numerical integrator can handle. The Von Neumann stability analysis [61] helps determine analogous CFL conditions for linear equations with arbitrary discretizations. The stability of numerical schemes for general nonlinear equations remains a topic of study. We refer to [268] for additional discussions on single-scale finite difference schemes. In general, the presence of a stiff coefficient ϵ^{-1} in equation (3.1) requires h and k to scale with ϵ in order to guarantee the stability of numerical integration schemes. This makes the numerical approximation of the solution of (3.1) computationally untractable when ϵ is close to 0.

3.1.2 Multiscale FLAVORization and general methodology

FLAVORs are multiscale in the sense that they accelerate computation by adopting both larger time and space steps. A finite difference scheme can be FLAVORized by employing two rules:

First, instead of a uniform mesh, use a mesh as depicted in Figure 3.1, in which a uniform spatial grid corresponds to a mesoscopic space step K that does not scale with ϵ, and an alternating temporal grid corresponds to two time steps, microscopic h (scaling with ϵ) and mesoscopic $H - h$ (H independent from ϵ). It is worth mentioning that when using this non-uniform mesh, grid sizes have to be taken into consideration when derivatives are approximated by finite differences. 1st-order derivatives are straightforward to obtain, and we refer to Section 3.2 for approximations of higher-order derivatives.

Second, the stiff parameter ϵ^{-1} should be temporarily set to be 0 (i.e., turned off) when the current time step is the mesoscopic $H - h$; if the small time step h is used instead, the large value of ϵ^{-1} needs to be restored, or in other words, stiffness should be turned on again.

The rule of thumb is that k and h should be chosen such that the integration of (3.1) with these step sizes and stiffness turned on is stable and accurate. On the other hand, there is another pair of step size values such that the same integration

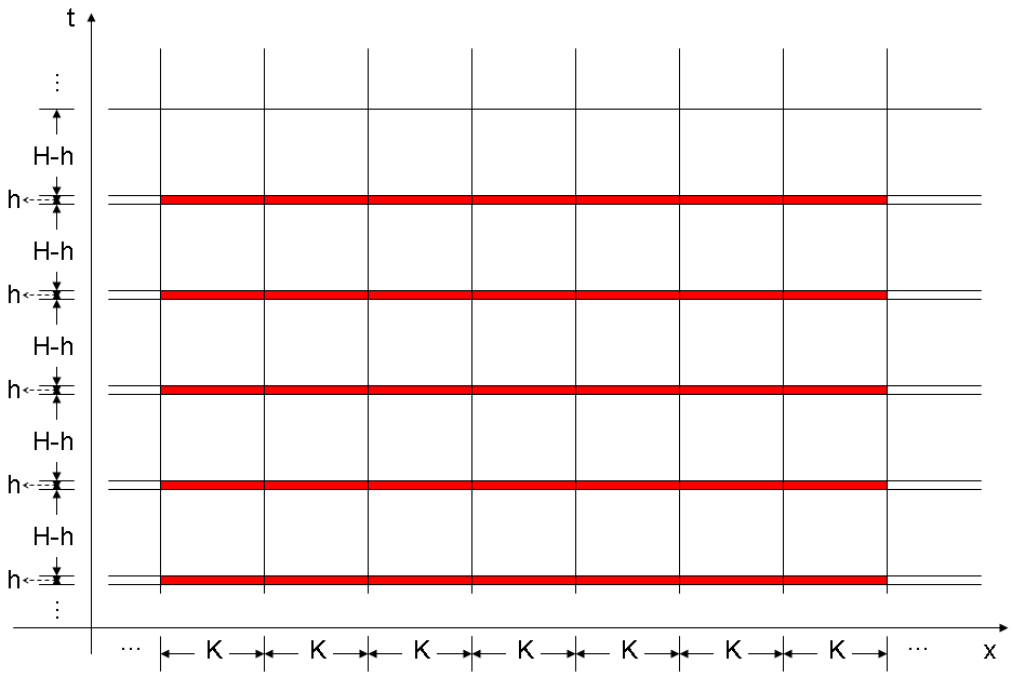


Figure 3.1: Mesh used by FLAVORs. A uniform mesoscopic space step is used and two alternating microscopic and mesoscopic time steps are used. Stiffness is turned on in red regions and turned off otherwise.

with stiffness turned off is stable and accurate, and K and H should be chosen to be an order of magnitude smaller than these values. FLAVORs does not require a microscopic k, but only a mesoscopic space-step K, a microscopic time-step h, and a mesoscopic time-step H.

The intuition is as follows: adopt the point of view of semi-discrete approach for PDE integration, in which space is discretized first and the PDE is approximated by a system of ODEs. The integration (in the time) of the resulting finite dimensional ODE system can be accelerated by applying the FLAVOR strategy to any legacy scheme (used as a black box). Turning on and off stiff coefficients in the legacy scheme and alternating microscopic time steps (stiffness on) with mesoscopic time steps (stiffness on) preserves the symmetries of that scheme and at the same time induces an averaging of the dynamic of (possibly hidden) slow variables with respect to the fast ones. With this strategy, the FLAVORized scheme

advances in mesoscopic time steps without losing stability. The (possibly hidden) slow dynamic is captured in a strong sense, while the fast one is captured only in the (weak) sense of measures. A rigorous proof of convergence of the proposed method relies on the assumption of existence of (possibly hidden) slow variables and of local ergodicity of (possibly hidden) fast variables (we refer to Section 3.4). It is important to observe that the proposed method does not require the identification of slow variables.

3.1.3 Example: Conservation law with Ginzburg-Landau source

Consider a specific stiff PDE:

$$u_t + f(u)_x = \epsilon^{-1} u(1 - u^2) \tag{3.2}$$

in which $f(u) = \sin u$ and $0 < \epsilon \ll 1$. Use the boundary condition of $u(x = 0, t) = u(x = L, t)$ and the initial condition of $u(x, t = 0) = \sin(\pi x)$. This system contains two scales: the fast process corresponds to u quickly converging towards 1 or -1, and the slow process corresponds to the front (with steep gradients) that separates $u > 0$ from $u < 0$ propagating at an $\mathcal{O}(1)$ velocity.

We will FLAVORize the following Lax-Friedrichs finite difference scheme:

$$\begin{cases} u_{i+1,j+1} & = \bar{u}_{i+1,j} - h\left(f_u(\bar{u}_{i+1,j})\frac{u_{i+2,j} - u_{i,j}}{2k} + \epsilon^{-1}\bar{u}_{i+1,j}(1 - \bar{u}_{i+1,j}^2)\right) \\ \bar{u}_{i+1,j} & \triangleq \frac{u_{i+2,j} + u_{i,j}}{2} \end{cases} \tag{3.3}$$

where $u_{i,j} = u_{i+L/k,j}$ and $u_{i,1} = \sin(\pi(i-1)k)$. If the domain of integration is restricted to $[0, L] \times [0, T]$, then $i = 1, 2, \ldots, \lfloor L/k \rfloor + 1$, and $j = 1, 2, \ldots, \lfloor T/h \rfloor + 1$. We use $h = 0.1\epsilon$ and $k = 0.2\epsilon$ for our purposes, both of which we found numerically at the order of the stability limit. In our experiment, we chose $\epsilon = 2 \cdot 10^{-3}$, and therefore $h = 0.0002$ and $k = 0.0004$.

The FLAVORized version of this scheme is:

$$\begin{cases} \tilde{u}_{i+1,j} & = \bar{u}_{i+1,j} - h\left(f_u(\bar{u}_{i+1,j})\frac{u_{i+2,j}-u_{i,j}}{2K} + \epsilon^{-1}\bar{u}_{i+1,j}(1-\bar{u}_{i+1,j}^2)\right) \\ \bar{u}_{i+1,j} & \triangleq (u_{i+2,j}+u_{i,j})/2 \\ u_{i+1,j+1} & = \frac{\tilde{u}_{i+2,j}+\tilde{u}_{i,j}}{2} - (H-h)\left(f_u(\frac{\tilde{u}_{i+2,j}+\tilde{u}_{i,j}}{2})\frac{\tilde{u}_{i+2,j}-\tilde{u}_{i,j}}{2K}\right) \end{cases} \tag{3.4}$$

where $u_{i,j} = u_{i+L/K,j}$ and $u_{i,1} = \sin(\pi(i-1)K)$. If the domain of integration is restricted to $[0,L]\times[0,T]$, then $i = 1, 2, \ldots, \lfloor L/K \rfloor + 1$, and $j = 1, 2, \ldots, \lfloor T/H \rfloor + 1$. We use the same h as before, and choose $H = 0.005$ and $K = 0.01$, which ensures that the stability of the integration remains independent of ϵ.

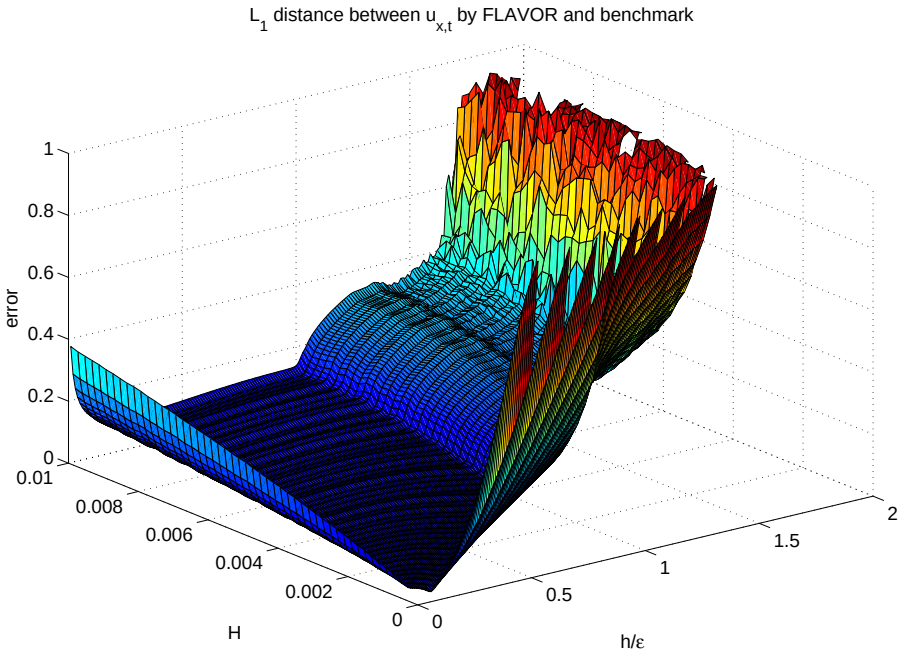


Figure 3.2: Errors of FLAVOR based on Lax-Friedrichs as a function of H and h. H samples multiples of 0.1ϵ, starting from 2x to 50x with 1x increment, and h ranges from 0.01ϵ to 3ϵ with 0.01ϵ increment. Errors with magnitude bigger than 1 are not plotted, for they indicate unstable integrations.

Errors of FLAVOR based on Lax-Friedrichs with different H and h values are computed by comparing the results to a benchmark Lax-Friedrichs integration with fine steps $h = 0.1\epsilon$ and $k = 0.2\epsilon$. More precisely, we calculated the distance

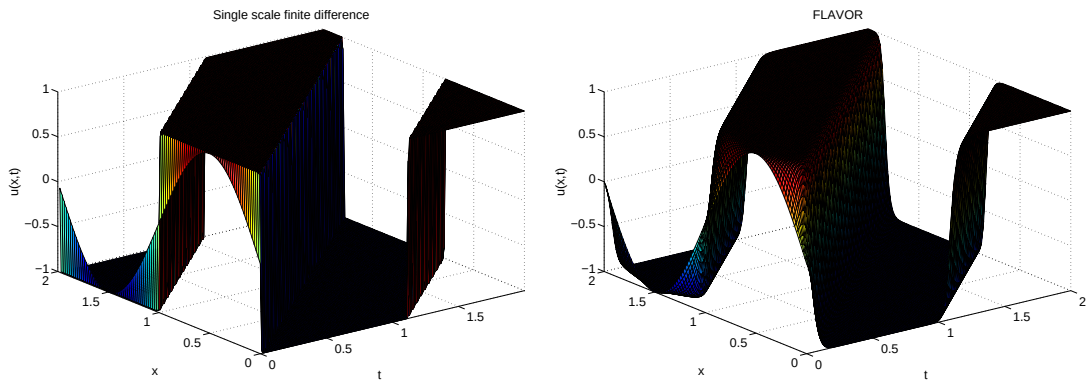


Figure 3.3: Numerical solutions to (3.2) by Lax-Friedrichs (left, (3.3)) and its FLAVORization (right, (3.4)).

between two vectors respectively corresponding to FLAVOR and Lax-Friedrichs integrations, which contain ordered $u(x,t)$ values on the intersection of FLAVOR and Lax-Friedrichs meshes (which is in fact the FLAVOR mesh as long as H is a multiple of 0.1ϵ). 1-norm is used and normalized by the number of discrete points to mimic the L^1 norm for the continuous solution. Experimental settings are $\epsilon = 2 \cdot 10^{-3}$, $L = 2$ and $T = 2$. As we can see in Figure 3.2, FLAVOR is indeed uniformly convergent in the sense that the error scales with H, as long as h takes an appropriate value. This is not surprising, because we have already proven in the ODE case that the error is bounded by a function of H (uniformly in ϵ) as long as $\left(\frac{h}{\epsilon}\right)^2 \ll H \ll h/\epsilon$, and this error can be made arbitrarily small as $H \downarrow 0$ (notice H can still be much larger than ϵ as $\epsilon \downarrow 0$).

Also, a typical run of FLAVOR ($H = 0.005$ and $K = 0.01$) in comparison to the benchmark ($h = 0.0002$ and $k = 0.0004$) is shown in Figure 3.3. FLAVOR captured the slow process strongly in the sense that it obtained the correct speeds of both steep gradients' propagations (up to arithmetic error and fringing). In this setting, FLAVOR achieves a $\frac{HK}{2hk} = 312.5$-fold acceleration. It is worth restating that both spatial and temporal step lengths of FLAVOR are mesocopic, whereas the counterparts in a single scale finite difference method have to be both microscopic for stability. The computational gain by FLAVOR will go to infinity as $\epsilon \to 0$, and

this statement will be true for all FLAVOR examples shown in this chapter.

3.2 Multisymplectic integrator for Hamiltonian PDEs

3.2.1 Single-scale method

We refer to [49, 193, 195] for a discussion on the geometry of Hamiltonian PDEs (e.g., multi-symplectic structure). We will now recall the Euclidean coordinate form of a Hamiltonian PDE:

$$\mathcal{M}z_t + \mathcal{K}z_x = \nabla_z H(z) \tag{3.5}$$

where $z(x,t)$ is a n-dimensional vector, $\mathcal{M}$ and $\mathcal{K}$ are arbitrary skew-symmetric matrices on $\mathbb{R}^n$, and $H : \mathbb{R}^n \to \mathbb{R}$ is an arbitrary smooth function. The solution preserves the multi-symplectic structure in the following sense:

$$\partial_t \iota(U,V) + \partial_x \kappa(U,V) = 0 \tag{3.6}$$

where ι and κ are differential 2-forms defined by

$$\iota(x,y) = \langle \mathcal{M}x, y\rangle \quad \text{and} \quad \kappa(x,y) = \langle \mathcal{K}x, y\rangle \tag{3.7}$$

and U and V are two arbitrary solutions to the variational equation (the solution is identified with $dz : \mathbb{R}^2 \mapsto \mathbb{R}^n$):

$$\mathcal{M}dz_t + \mathcal{K}dz_x = D_{zz}H(z)dz, \quad dz(x,t) \in \mathbb{R}^n \tag{3.8}$$

Preservation of multi-symplecticity can be partially and intuitively interpreted as a conservation of infinitesimal volume in the jet bundle, which generalizes the conservation of phase space volume in Hamiltonian ODE settings to field theories.

A broad spectrum of PDEs fall in the class of Hamiltonian PDEs, including generalized KdV, nonlinear Schrödinger models, nonlinear wave equations, atmo-

spheric flows, fluid-structure interactions, etc. [46, 45, 48, 49]. We also refer to [50] and references therein for surveys on numerical recipes, and to [183] for an application to numerical nonlinear elastodynamics.

Hamiltonian PDEs (3.5) can be viewed as Euler-Lagrange equations for field theories, which are obtained by applying Hamilton's principle (i.e., a variational principle of $\delta\mathcal{S}/\delta z = 0$) to the following action:

$$\mathcal{S}(z(\cdot,\cdot)) = \iint \mathcal{L}(z, z_t, z_x)\, dt\, dx \tag{3.9}$$

where the Lagrangian density is given by

$$\mathcal{L}(z, z_t, z_x) = \frac{1}{2}\langle \mathcal{M} z_t, z\rangle + \frac{1}{2}\langle \mathcal{K} z_x, z\rangle - H(z) \tag{3.10}$$

This variational view of Hamiltonian PDEs will intrinsically guarantee the preservation of multi-symplecticity, and there will be a field generalization of Noether's theorem, which ensures conservation of momentum maps corresponding to symmetries.

Numerically, instead of discretizing the equation (3.5), we prefer the approach of variational integrators because they are intrinsically multi-symplectic and therefore structure-preserving [193, 195, 192, 183]. These integrators are obtained as follows: first discretize the action (3.9) using quadratures, then apply variational principle to the discrete action (which depends on finitely many arguments), and finally, solve the algebraic system obtained from the variational principle, i.e., the discrete Euler-Lagrange equations.

For an illustration, consider a nonlinear wave equation:

$$u_{tt} - u_{xx} = V'(u) \tag{3.11}$$

with periodic boundary condition $u(x + L, t) = u(x, t)$ and compatible initial conditions $u(x, t = 0) = f(x)$ and $u_t(x, t = 0) = g(x)$. Suppose we are interested in the solution in a domain $[0, L] \times [0, T]$.

Rewrite the high-order PDE as a system of first-order PDEs (notice these covariant equations can be obtained through an intrinsic procedure, which works on manifolds as well [47]):

$$v_t - w_x = V'(u) \tag{3.12}$$

$$u_t = v \tag{3.13}$$

$$u_x = w \tag{3.14}$$

The corresponding Lagrangian density is:

$$\mathcal{L} = \frac{1}{2}u_t^2 - \frac{1}{2}u_x^2 + V(u) \tag{3.15}$$

Using a forward time forward space approximation, we obtain the following discrete Lagrangian:

$$L_{i,j}^d \triangleq h_{ij}k_{ij}\left[\frac{1}{2}\left(\frac{u_{i,j+1}-u_{i,j}}{h_{ij}}\right)^2 - \frac{1}{2}\left(\frac{u_{i+1,j}-u_{i,j}}{k_{ij}}\right)^2 + V(u_{i,j})\right] \tag{3.16}$$

$$\approx \int_{t_j}^{t_{j+1}=t_j+h_{ij}} dt \int_{x_i}^{x_{i+1}=x_i+k_{ij}} dx \left[\frac{1}{2}u_t^2 - \frac{1}{2}u_x^2 + V(u)\right] \tag{3.17}$$

where space step k_{ij} and time step h_{ij} define a rectangular grid of size $k_{ij} \times h_{ij}$. The simplest single-scale choice would be $k_{ij} = k$ and $h_{ij} = h$ for some k and h.

As a consequence, the continuous action $\mathcal{S}$ is approximated by a discrete action:

$$\mathcal{S}_d = \sum_{\alpha=1}^{N}\sum_{\beta=1}^{M} L_{\alpha,\beta}^d \approx \mathcal{S} = \iint \mathcal{L}\, dt\, dx \tag{3.18}$$

and Hamilton's principle of least action $\delta\mathcal{S}_d = 0$ gives

$$\frac{\partial}{\partial u_{i,j}} \sum_{\alpha=1}^{N}\sum_{\beta=1}^{M} L_{\alpha,\beta}^d = 0 \tag{3.19}$$

for $1 \le i \le N$ and $1 \le j \le M$, where N and M are such that $\sum_{\alpha=1}^{N} k_{\alpha\beta} = L$ for any β and $\sum_{\beta=1}^{M} h_{\alpha\beta} = T$ for any α.

Taking derivative with respect to $u_{i,j}$, we obtain the following discrete Euler-Lagrange equations:

$$k_{ij}\frac{u_{i,j}-u_{i,j+1}}{h_{ij}}-h_{ij}\frac{u_{i,j}-u_{i+1,j}}{k_{ij}}+h_{ij}k_{ij}V'(u_{i,j})+k_{i,j-1}\frac{u_{i,j}-u_{i,j-1}}{h_{i,j-1}}-h_{i-1,j}\frac{u_{i,j}-u_{i-1,j}}{k_{i-1,j}}=0 \tag{3.20}$$

The system of above equations is explicitly solvable when equipped with boundary conditions and initial conditions; for instance, below is a consistent discretization of the continuous version:

$$\begin{cases} u_{i,j}=u_{i+N,j}, & \forall i,j \\ u_{i,1}=f\left(\sum_{\alpha=1}^{i}k_{\alpha 1}\right), & \forall i \\ u_{i,2}=u_{i,1}+h_{i1}g\left(\sum_{\alpha=1}^{i}k_{\alpha 2}\right), & \forall i \end{cases} \tag{3.21}$$

This numerical recipe is convergent. In fact, multi-symplectic integrators obtained from variational principles can be viewed as special members of finite difference methods, whose error analysis is classical.

It is worth pointing out that the above procedure works for any Hamiltonian PDEs of form (3.5). Also, notice that high-order derivatives are dealt with in an intrinsic way regardless of whether the mesh is uniform.

3.2.2 FLAVORization of multi-symplectic integrators

Now consider a multiscale Hamiltonian PDE

$$\mathcal{M}(1,\epsilon^{-1})z_t+\mathcal{K}(1,\epsilon^{-1})z_x=\nabla_z H(1,\epsilon^{-1},z) \tag{3.22}$$

Any single-scale multi-symplectic integrator can be FLAVORized (to achieve computational acceleration) by using the following strategy: (i) Use the two-scale mesh illustrated in Figure 3.1, and (ii) turn off large coefficients when taking mesoscopic time-steps. Unlike FLAVORizing a general finite difference scheme,

we FLAVORize the action $\mathcal{S}_d$ instead of the PDE. Specifically, choose

$$\begin{cases} k_{ij} = K, & \forall i, j \\ h_{ij} = h, & \forall i \text{ and odd } j \\ h_{ij} = H - h, & \forall i \text{ and even } j \end{cases} \tag{3.23}$$

and let $\epsilon^{-1} = 0$ in $L^d_{i,j}$ for even j's and all i's, while the large value of ϵ^{-1} is kept in $L^d_{i,j}$ for odd j's and all i's. h and H correspond to a microscopic and a mesoscopic time-step, and K corresponds to a mesoscopic space-step; the same rule of thumb for choosing them in Section 3.1 applies.

After applying the discrete Hamilton's principle, the resulting discrete Euler-Lagrange equations corresponding to a multi-symplectic integrator will still be (3.20), except that stiffness is turned off in half of the grids. Multisymplecticity is automatically gained, because the updating equations originate from a discrete variational principle [193].

3.2.3 Example: Multiscale Sine-Gordon wave equation

Consider a specific nonlinear wave equation (3.11) in which $V(u) = -\cos(\omega u) - \cos(u)$. If $\omega = 0$, this corresponds to the Sine-Gordon equation, which has been studied extensively due to its soliton solutions and its relationships with quantum physics (for instance, as a nonlinear version of Klein-Gordon equation). We are interested in the case in which ω (identified with ϵ^{-1} in this case) is big, so that a separation of timescale exhibits.

Arbitrarily choose $L = 2$ and use periodic boundary condition $u(x + L, t) = u(x, t)$, and let initial condition be $u(x, 0) = \sin(2\pi x/L)$ and $u_t(x, 0) = 0$. Denote total simulation time by T. Use the FLAVOR mesh (3.23). In order to obtain a stable and accurate numerical solution, k and h have to be $o(1/\omega)$, and K and H need to be $o(1)$.

A comparison between the benchmark of the single-scale forward time forward space multi-symplectic integrator ((3.20) with $h_{ij} = h$ and $k_{ij} = k$) and its FLA-

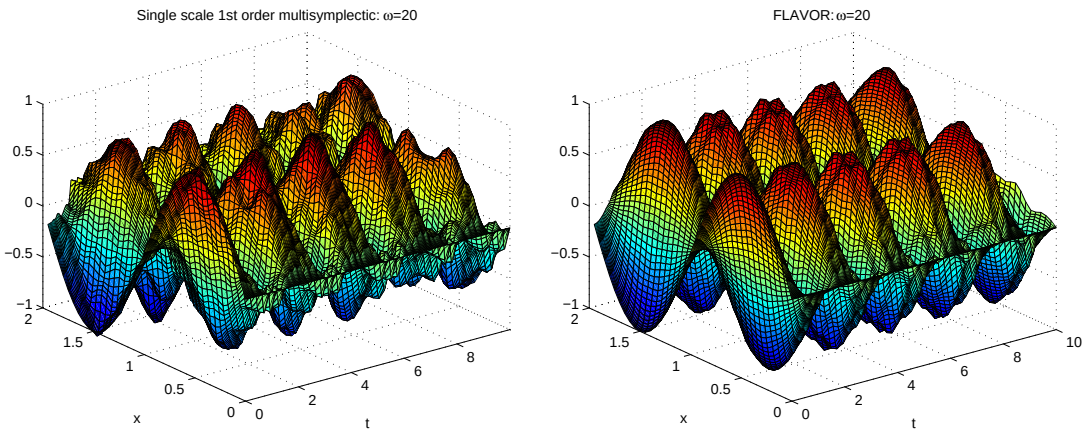


Figure 3.4: Numerical solutions to multiscale Sine-Gordon equation by single-scale 1st-order multi-symplectic integrator (left) and its FLAVORization (right). For clarity, the surface plots (but not simulations) use the same mesh size.

VORization ((3.20) with mesh (3.23) and $V'(u) = \omega \sin(\omega u) + \sin(u)$ for odd j and $V'(u) = \sin(u)$ for even j) is presented in Figure 3.4. $\omega = 20$, $k = L/20/\omega$ and $h = k/2$, and $K = L/40$ and $H = K/2$. It is intuitive to say that the slow process of wave propagation is well-approximated by FLAVOR, although the fast process of local fluctuation is not captured in the strong sense. Error quantification is not done, because what the slow and fast processes are is not rigorously known here. $HK/2hk = 50$-fold acceleration is obtained by FLAVOR.

Readers familiar with the splitting theory of ODEs [199] might question whether FLAVORs are equivalent to an averaged stiffness of $\tilde{\omega} = \omega \frac{h}{H}$ (which corresponds $\tilde{\omega} = 2$ in the numerical experiment described above). The answer is no, because the equivalency given by the splitting theory is only local. In fact, the same single-scale forward time forward space multi-symplectic integration of the case $\omega = 2$ is shown in Figure 3.5, which is clearly distinct from the FLAVOR result in Figure 3.4.

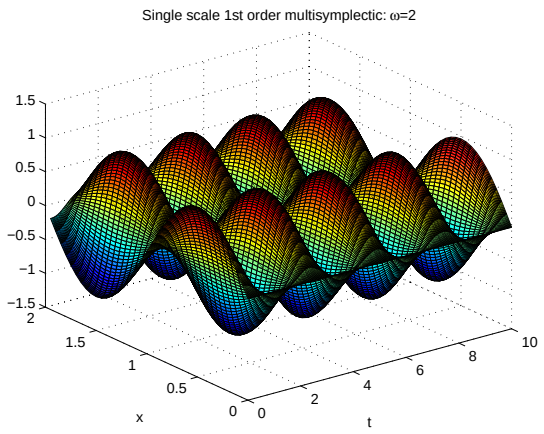


Figure 3.5: Numerical solutions to multiscale Sine-Gordon equation with the 'equivalent' stiffness by single-scale 1st-order multi-symplectic integrator. For clarity, the surface plot (but not the simulation) uses the same mesh size (as in Figure 3.4).

3.3 Pseudospectral methods

3.3.1 Single-scale method

Consider a PDE

$$u_t(x,t) = \mathcal{L}u(x,t) \tag{3.24}$$

with periodic boundary condition $u(x,t) = u(x+L,t)$ and initial condition $u(x,0) = f(x)$, where $\mathcal{L}$ is a differential operator involving only spatial derivatives.

The Fourier collocation method approximates the solutions by the truncated Fourier series:

$$u_N(x,t) = \sum_{|n|\leq N/2} a_n(t) e^{in2\pi x/L} \tag{3.25}$$

and solves for $a_n(t)$'s by requiring the PDE to hold at collocation points y_j:

$$\partial_t u_N(y_j,t) - \mathcal{L}u_N(y_j,t) = 0 \tag{3.26}$$

This yields a system of N ODEs, which can be integrated by any favorite ODE solver. Of course, specific choices of collocations points will affect the numerical approximation. Oftentimes, the simplest choice of $y_j = Lj/N, j = 0,\ldots,N-1$ is

used, and in this case, the method is also called a pseudospectral method. We refer to [139] for additional details on Fourier collocation methods. It is worth mentioning that pseudospectral methods can also be multi-symplectic when applied to Hamiltonian PDEs [62].

3.3.2 FLAVORization of pseudospectral methods

When the PDE is stiff (for instance, when $\mathcal{L}$ contains a large parameter ϵ^{-1}), FLAVORs can be employed to integrate the stiff ODEs (which will still contain ϵ^{-1}) resulting from a pseudospectral discretization.

Similarly, for the FLAVORization of a pseudospectral method, it is sufficient to choose $N \gg L$ instead of $N \gg \epsilon^{-1}L$, i.e., the space-step can be coarse ($K = o(1)$). For time stepping, alternatively switching between $h = o(\epsilon)$ and $H - h$ for a mesoscopic $H = o(1)$ is again needed, and stiffness has to be turned off over the mesoscopic step of $H - h$. In a sense, we are still using the same FLAVOR 'mesh' (Figure 3.1), except that here we do not discretize space, but instead truncate Fourier series to resolve the same spatial grid size.

3.3.3 Example: A slow process driven by a non-Dirac fast process

Consider the following system of PDEs

$$\begin{cases} u_t + u_x - q^2 = 0 \\ q_t + q_x - p = 0 \\ p_t + p_x + \omega^2 q = 0 \end{cases} \tag{3.27}$$

with periodic boundary conditions $u(x,t) = u(x+L,t)$, $q(x,t) = q(x+L,t)$, and $p(x,t) = p(x+L,t)$, and initial conditions $u(x,0) = f^u(x)$, $q(x,0) = f^q(x)$, and $p(x,0) = f^p(x)$. The integration domain is restricted to $[0,T] \times [0,L]$. The stiffness ϵ^{-1} is identified with ω^2. We choose the initial condition of $f^u(x) = f^q(x) = \cos(2\pi x/L)$ and $f^p(x) = 0$.

In this system, q and p correspond to a fast process, which is a field theory

version of a harmonic oscillator with high frequency ω. u is a slow process, into which energy is pumped by the fast process in a nontrivial way.

We have chosen to FLAVORize (3.27) because it does not fall into the (simpler) category of systems with fast processes converging towards Dirac (single point support) invariant distributions [102].

We use the classical 4th-order Runga-Kutta scheme (see, for instance, [129]) for the (single-step) time integration of the pseudospectrally discretized system of ODEs (3.26). Write $\phi_h^{\omega^2} : \tilde{a}_n^{u,q,p}(t) \mapsto \tilde{a}_n^{u,p,q}(t+h)$ its numerical flow over a microscopic time step h (consisting of four sub-steps), where $\tilde{a}_n^{u,q,p}(t)$ are numerical approximations to the Fourier coefficients in (3.25), for the unknowns u, q and p at an arbitrary time t. Then, the corresponding FLAVOR update over a mesoscopic time step H will be $\phi_{H-h}^0 \circ \phi_h^{\omega^2}$, which consists of eight sub-steps.

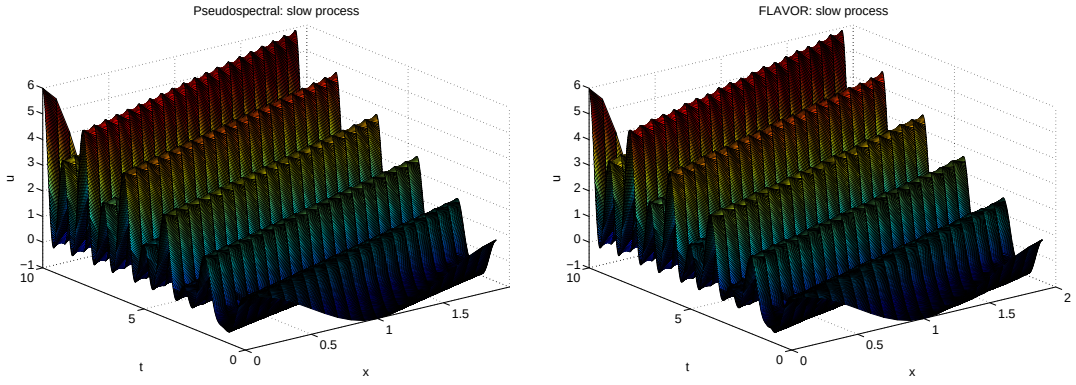


Figure 3.6: Single-scale (left) and multiscale pseudospectral (right) integrations of slow u in system (3.27). Plotting mesh for the single-scale simulation is coarser than its computation mesh.

We present in Figure 3.6 and Figure 3.7 a comparison between the benchmark of single-scale pseudospectral simulation and its FLAVORization. It can be seen that the slow process of u is captured in strong (point-wise) sense, whereas the fast process of q is only approximated in a weak sense (i.e., as a measure, in the case wave shape and amplitude are correct, but not the period). We choose $L = 2$, $T = 10$ and $\omega = 1000$. The single-step integration uses $N = 20$ and $h = 0.1/\omega$ (notice that this is already beyond the stability/accuracy region of a

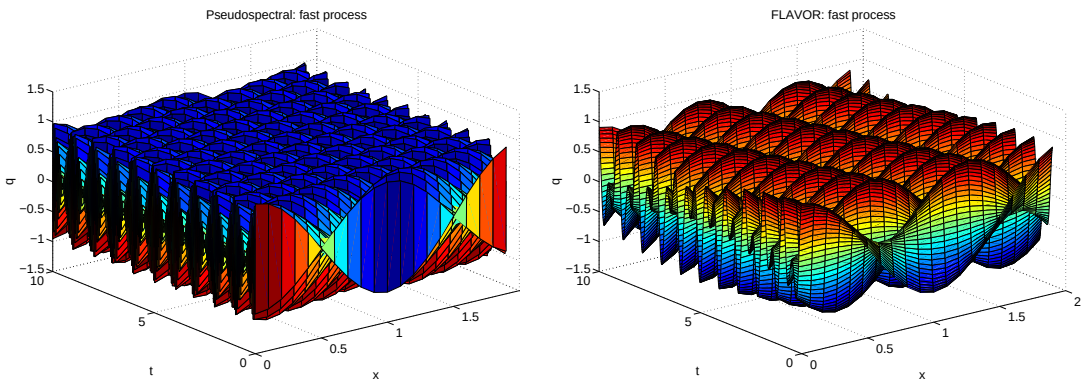


Figure 3.7: Single-scale (left) and multiscale pseudospectral (right) integrations of fast q in system (3.27). Plotting mesh for the single-scale simulation is coarser than its computation mesh. The same color does not indicate the same value in these two plots.

single-scale finite difference, since the space step does not depend on $1/\omega$; the spectral method is more stable/accurate for a large space-step), and FLAVOR uses $N = 20$, $h = 1/\omega^2$ and $H = 0.01$. $H/2h = 50$-fold acceleration is achieved by FLAVOR.

3.4 Convergence analysis

3.4.1 Semi-discrete system

All FLow AVeraging integratORS described in previous sections are illustrations of the following (semi-discrete) strategy: first, space is discretized or interpolated; next, spatial differential operators are approximated by algebraic functions of finitely many spatial variables; finally, the resulting system of ODEs is numerically integrated by a corresponding ODE-FLAVOR (see Chapter 2 or [274]). In this section, we will use the semi-discrete ODE system as an intermediate link to demonstrate that these PDE-FLAVORs are convergent to the exact PDE solution under reasonable assumptions (in a strong sense with respect to (possibly hidden) slow variables and in the sense of measures with respect to fast variables).

More precisely, consider a spatial mesh (vector) $\mathcal{M}^S = [x_1, x_2, \ldots]$, a temporal mesh (vector) $\mathcal{M}^T = [t_1, t_2, \ldots]$, and a domain mesh (matrix) $\mathcal{M} = \mathcal{M}^S \times \mathcal{M}^T$.

Examples of these meshes include the FLAVOR mesh $\mathcal{M}^S = [K, 2K, \ldots, NK]$ and $\mathcal{M}^T = [h, H, H+h, 2H, \ldots, (M-1)H, (M-1)H+h, MH]$, and a usual single-scale (step) integration mesh $\mathcal{M}^S = [k, 2k, \ldots, L]$ and $\mathcal{M}^T = [h, 2h, \ldots, T]$ (recall the domain size is $L = NK$ by $T = MH$). We will use the FLAVOR mesh throughout this section. We will compare the solution of the PDE (3.28) with the solution obtained with the FLAVOR strategy at these discrete points.

For simplicity, assume the PDE of interest is 1st-order in time derivative:

$$u_t(x,t) = F(1, \epsilon^{-1}, x, t, u(x,t), u_x(x,t), \ldots) \tag{3.28}$$

Observe that a PDE (3.1) with higher-order time derivatives can be written as a system of 1st-order (in time derivatives) PDEs.

Now consider a consistent discretization of PDE (3.28) with space step K and time step h (we refer to Page 20 of [268] for a definition of the notion of consistency, which intuitively means vanishing local truncation error). Letting $h \downarrow 0$ in this discretization, we obtain a semi-discrete system (continuous in time and discrete in space). This semi-discrete system is denoted by the following system of ODEs, with approximated spatial derivatives:

$$\begin{cases} \dot{u}_1(t) = f_1(u_1, u_2, \ldots, u_N, \epsilon^{-1}, t) \\ \dot{u}_2(t) = f_2(u_1, u_2, \ldots, u_N, \epsilon^{-1}, t) \\ \quad \ldots \\ \dot{u}_N(t) = f_N(u_1, u_2, \ldots, u_N, \epsilon^{-1}, t) \end{cases} \tag{3.29}$$

Assuming existence and uniqueness of an exact $\mathcal{C}^1$ strong solution u to the PDE (3.28), and writing $u(\mathcal{M}_i^S, t)$ its values at the spatial discretization points, we define for each i the following remainder:

$$\mathcal{R}_i(\epsilon^{-1}, t) \triangleq \frac{\partial u}{\partial t}(\mathcal{M}_i^S, t) - f_i(u(\mathcal{M}_1^S, t), u(\mathcal{M}_2^S, t), \ldots, u(\mathcal{M}_N^S, t), \epsilon^{-1}, t) \tag{3.30}$$

which is a real function of t indexed by ϵ^{-1}.

Then, $u_i(t)$ approximates the exact solution $u(\mathcal{M}_i^S, t)$ evaluated at grid points in the sense that these remainders vanish as $\epsilon^{-1}K \downarrow 0$ (where $K := \mathcal{M}_i^S - \mathcal{M}_{i-1}^S$):

Lemma 3.4.1. *Assume that F in (3.28) satisfies*

$$|F(1, \epsilon^{-1}, x, t, u(x,t), u_x(x,t), \ldots)| \leq (1+\epsilon^{-1})|F(1, 1, x, t, u(x,t), u_x(x,t), \ldots)| \tag{3.31}$$

Assume that the f_i in (3.29) satisfies similar inequalities. Then, there exists a constant C_i independent from ϵ, h, H or K, such that for bounded t and u

$$|\mathcal{R}_i(\epsilon^{-1}, t)| \leq (1+\epsilon^{-1})C_i K \tag{3.32}$$

Remark 3.4.1. (3.31) *is true, for instance, in cases where*

$$F(1, \epsilon^{-1}, x, t, u(x,t), \ldots) = F_0(x, t, u(x,t), \ldots) + \epsilon^{-1}F_1(x, t, u(x,t), \ldots). \tag{3.33}$$

Proof. The linear scaling with K in (3.32) immediately follows from the definition of consistency, and the parameter $1 + \epsilon^{-1}$ in (3.32) has its origin in (3.31). □

Remark 3.4.2. *The consistency of finite difference methods can be easily shown using Taylor expansions. For instance, applying a Taylor expansion to the solution of $u_t - \epsilon^{-1}u_x = a(u)$ leads to*

$$\begin{aligned} u(iK, (j+1)h) =& u(iK, jh) + h\Big(\epsilon^{-1}\big(\frac{u((i+1)K, jh) - u(iK, jh)}{K} \\ &+ \mathcal{O}(K)\big) + a(u(iK, jh))\Big) + \mathcal{O}(h^2) \end{aligned} \tag{3.34}$$

which implies

$$\frac{\partial}{\partial t}u(iK, t) = \epsilon^{-1}\frac{u((i+1)K, t) - u(iK, t)}{K} + a(u(iK, t)) + \epsilon^{-1}\mathcal{O}(K) \tag{3.35}$$

and naturally establishes the correspondence of $f_i(u_1, \ldots, u_N, \epsilon^{-1}, t) = \epsilon^{-1}\frac{u_{i+1}(t) - u_i(t)}{K} + a(u_i(t))$ and $\mathcal{R}_i = \epsilon^{-1}\mathcal{O}(K)$ for a 1st-order finite difference scheme. Notice that

the remainders are still stiff, but we will see later that this is not a problem, since they can be handled by ODE-FLAVORs. The consistency of pseudospectral method can be shown similarly using Fourier analysis.

With $\mathcal{R}_i$ defined in (3.30), consider the following system of ODEs:

$$\begin{cases} \dot{u}_1(t) = f_1(u_1, u_2, \ldots, u_N, \epsilon^{-1}, t) + \mathcal{R}_1(\epsilon^{-1}, t) \\ \quad \ldots \\ \dot{u}_N(t) = f_N(u_1, u_2, \ldots, u_N, \epsilon^{-1}, t) + \mathcal{R}_N(\epsilon^{-1}, t) \end{cases} \tag{3.36}$$

with initial condition $u_i(0) = u(\mathcal{M}_i^S, 0)$. Obviously, its solution $(u_i(t))_{1\le i\le N}$ is the exact PDE solution sampled at spatial grid points, i.e., $u_i(t) = u(\mathcal{M}_i^S, t)$.

We will now establish the accuracy of PDE-FLAVOR by showing that an ODE-FLAVOR integration of (3.36) leads to an accurate approximation of $(u_i(t))_{1\le i\le N}$. Since space (with fixed width L) is discretized by N grid points, we use the following (normalized by N) norm in our following discussion (suppose $v_i(t) = v(\mathcal{M}_i^S, t)$ for a function v):

$$\|[v_1(t), v_2(t), \ldots, v_N(t)]\| \triangleq \frac{1}{N} \|[v_1(t), v_2(t), \ldots, v_N(t)]\|_1 \tag{3.37}$$

Observe that if $v(\cdot, t)$ is Riemann integrable, then

$$\lim_{K\downarrow 0} \left\|[v(\mathcal{M}_1^S, t), v(\mathcal{M}_2^S, t), \ldots, v(\mathcal{M}_N^S, t)]\right\| \to \frac{1}{L} \|v(\cdot, t)\|_{\mathcal{L}^1} \qquad \text{(recall } L = NK \text{ is fixed)}, \tag{3.38}$$

and hence the norm (3.37) does not blow up or vanish as $N \to \infty$.

3.4.2 Sufficient conditions and the two-scale convergence of PDE-FLAVORs

We will now prove the accuracy of PDE-FLAVORs under the assumption of existence of (possibly hidden) slow and locally ergodic fast variables. The convergence of PDE-FLAVORs will be expressed using the notion of two-scale flow convergence

introduced in Chapter 2 or [274], which corresponds to a strong convergence on the slow variables and a weak convergence on the fast ones.

The scale separation and local ergodicity assumption is analogous to Conditions 2.1.1 and 2.1.2, except that now we allow the scale separation diffeomorphism and the local ergodic measure to be slowly time-dependent:

Condition 3.4.1. *Assume that the ODE system* (3.36) *satisfies the following conditions:*

1. *(Existence of hidden slow and fast variables): There exists a (possibly time-dependent) diffeomorphism* $\eta^t : [u_1(t), \dots, u_N(t)] \mapsto [x(t), y(t)]$ *from* $\mathbb{R}^N$ *onto* $\mathbb{R}^{N-p} \times \mathbb{R}^p$ *with uniformly bounded* C^1, C^2 *derivatives with respect to* u_i*'s and* t*, and such that for all* $\epsilon > 0$*,* $(x(t), y(t))$ *satisfies*

$$\begin{cases} \dot{x}(t) &= f(x(t), y(t), t) \\ \dot{y}(t) &= \epsilon^{-1} g(x(t), y(t), t) \end{cases}, \tag{3.39}$$

   where f *and* g *have bounded* C^1 *derivatives with respect to* x*,* y *and* t*.*

2. *(Local ergodicity of vast variables): There exists a family of probability measures* $\mu^t(x, dy)$ *on* $\mathbb{R}^p$ *indexed by* $x \in \mathbb{R}^{N-p}$ *and* $t \in \mathbb{R}$*, and a family of positive functions* $T \mapsto E^t(T)$ *satisfying* $\lim_{T\to\infty} E^t(T) = 0$ *for all bounded* t*, such that for all* x_0, y_0, t_0, T *bounded and* ϕ *uniformly bounded and Lipschitz, the solution to*

$$\dot{Y}_t = g(x_0, Y_t, t_0) \qquad Y_0 = y_0 \tag{3.40}$$

   satisfies

$$\Big|\frac{1}{T}\int_0^T \phi(Y_s)ds - \int_{\mathbb{R}^p} \phi(y)\mu^{t_0}(x_0, dy)\Big| \le \chi^{t_0}\big(\|(x_0, y_0)\|\big) E^{t_0}(T)(\|\phi\|_{L^\infty} + \|\nabla\phi\|_{L^\infty}) \tag{3.41}$$

   where $r \mapsto \chi^{t_0}(r)$ *is bounded on compact sets, and* μ^t *has bounded derivative with respect to* t *in total variation norm.*

Under Conditions 3.4.1, the computation of the solution of PDE (3.28) can be accelerated by applying the FLAVOR strategy to a single-scale time integration of the semi-discretized system (3.29). We recap the construction of FLAVORs:

Write $\Phi^{\alpha}_{t,t+\tau}$ the numerical flow of a given (legacy) ODE integrator for (3.29):

$$\Phi^{\alpha}_{t,t+\tau} : [\tilde{u}_1(t), \ldots, \tilde{u}_N(t)] \mapsto [\tilde{u}_1(t+\tau), \ldots, \tilde{u}_N(t+\tau)]\,, \tag{3.42}$$

where $\tilde{u}_i(s)$ approximates $u_i(s)$ for all s, τ is the integration time step, and α is a controllable parameter that replaces the stiff parameter ϵ^{-1} in (3.29) and takes values of ϵ^{-1} (stiffness 'on') or 0 (stiffness 'off').

Definition 3.4.1 (ODE-FLAVORs)**.** *The FLow AVeraging integratOR associated with Φ is defined as the algorithm simulating the process:*

$$\begin{aligned}[\bar{u}_1(t), \ldots, \bar{u}_N(t)] = \big(\Phi^0_{(k-1)H+h,kH} \circ \Phi^{\frac{1}{\epsilon}}_{(k-1)H,(k-1)H+h}\big) \circ \cdots \\ \circ \big(\Phi^0_{H+h,2H} \circ \Phi^{\frac{1}{\epsilon}}_{H,H+h}\big) \circ \big(\Phi^0_{h,H} \circ \Phi^{\frac{1}{\epsilon}}_{0,h}\big)([u_1(0), \ldots, u_N(0)])\end{aligned} \tag{3.43}$$

where (the number of steps) k is a piece-wise constant function of t satisfying $kH \le t < (k+1)H$, h is a microscopic time step resolving the fast timescale ($h \ll \epsilon$), H is a mesoscopic time step independent of the fast timescale satisfying $h \ll \epsilon \ll H \ll 1$ and

$$(\frac{h}{\epsilon})^2 \ll H \ll \frac{h}{\epsilon} \tag{3.44}$$

In order for a FLAVOR to be convergent, the legacy code, of course, has to be consistent:

Condition 3.4.2. *Consider the legacy ODE integrator with one-step update map $\Phi^{\alpha}_{t,t+\tau}$ introduced in* (3.42)*. Suppose there exist constants $C > 0$ and $H_0 > 0$ independent of N and α, such that for any $\tau \le H_0 \min(1/\alpha, 1)$ and bounded vector*

$[u_1, \ldots, u_N]$,

$$\|\Phi^{\alpha}_{t,t+\tau}(u_1, \ldots, u_N) - [u_1, \ldots, u_N] - \tau[f_1(u_1, \ldots, u_N, \alpha, t), \ldots \\ \ldots, f_N(u_1, \ldots, u_N, \alpha, t)]\| \leq C\tau^2(1+\alpha)^2\,, \tag{3.45}$$

Observe that we are integrating (3.29) but not (3.36), since the remainders $\mathcal{R}_i$'s are a priori unknown unless the exact PDE solution is known. However, the following lemma implies that the FLAVORization of this integration is in fact convergent to the solution of (3.36), even though $\mathcal{R}_i$'s are possibly stiff. This is due to the joint effect of the FLAVOR mesh and the turned-on-and-off stiffness.

Lemma 3.4.2. *Assume that $\Phi^{\alpha}_{t,t+\tau}$, introduced in* (3.42)*, satisfies Condition 3.4.2. Let h and H be the time steps used in the FLAVORization 3.4.1. If $h \ll \epsilon$, $H \ll h/\epsilon$, and $K = \mathcal{O}(H)$, then*

$$\|\Phi^{\alpha}_{t,t+\tau}(u_1, \ldots, u_N) - [u_1, \ldots, u_N] - \tau[f_1(u_1, \ldots, u_N, \alpha, t) + \mathcal{R}_1(\alpha, t), \ldots \\ \ldots, f_N(u_1, \ldots, u_N, \alpha, t) + \mathcal{R}_N(\alpha, t)]\| \leq C\tau^2(1+\alpha)^2 \tag{3.46}$$

where $\tau = h$ when $\alpha = \epsilon^{-1}$ and $\tau = H - h$ when $\alpha = 0$.

Proof. By Condition 3.4.2, we have

$$\|\Phi^{\alpha}_{t,t+\tau}(u_1, \ldots, u_N) - [u_1, \ldots, u_N] - \tau[f_1(u_1, \ldots, u_N, \alpha, t), \ldots \\ \ldots, f_N(u_1, \ldots, u_N, \alpha, t)]\| \leq C\tau^2(1+\alpha)^2 \tag{3.47}$$

for any $\tau \leq \min(1/\alpha, 1)H_0$. In addition, Lemma 3.4.1 gives a bound on the remainders: when $\alpha = \epsilon^{-1}$, there exists a constant $\tilde{C} > 0$ independent of N and ϵ^{-1}, such that for all i,

$$|\tau\mathcal{R}_i(\epsilon^{-1}, t)| \leq \tau\tilde{C}K\epsilon^{-1} \tag{3.48}$$

Because we use $\tau = h$ in this case and $K \ll \epsilon^{-1}\tau$, the above is bounded by $\tau\tilde{C}(\hat{C}\epsilon^{-1}\tau)\epsilon^{-1} \leq C\tau^2(1+\alpha)^2$ for some constants $\hat{C} \ll 1$ and $C = \tilde{C}\hat{C}$. When

$\alpha = 0$ on the other hand, there exists a constant $\tilde{C} > 0$ such that for all i

$$|\tau \mathcal{R}_i(\epsilon^{-1}, t)| \leq \tau \tilde{C} K \tag{3.49}$$

Because $K = \mathcal{O}(H)$ and we use $\tau = H - h = \mathcal{O}(H)$ in this case, the above is bounded by $\tau \tilde{C} \hat{C} \tau \leq C\tau^2(1+\alpha)^2$ for some constants $\hat{C}$ and we let $C = \tilde{C}\hat{C}$. Notice that the value of K is fixed in both cases but τ has different values: the flow map used in FLAVOR associated with $\alpha = 0$ is the one with mesoscopic step $\Phi^0_{t+h,t+H}$, i.e., $\tau = H - h$; when $\alpha = \epsilon^{-1}$ on the other hand, the flow map is $\Phi^{\epsilon^{-1}}_{t,t+h}$ and $\tau = h$. Finally, the triangle inequality gives

$$\begin{aligned}
&\|\Phi^{\alpha}_{t,t+\tau}(u_1, \ldots, u_N) - [u_1, \ldots, u_N] - \tau[f_1(u_1, \ldots, u_N, \alpha, t) + \mathcal{R}_1(\alpha, t), \ldots \\
&\ldots, f_N(u_1, \ldots, u_N, \alpha, t) + \mathcal{R}_N(\alpha, t)]\| \leq \|\Phi^{\alpha}_{t,t+\tau}(u_1, \ldots, u_N) - [u_1, \ldots, u_N] - \\
&\tau[f_1(u_1, \ldots, u_N, \alpha, t), \ldots, f_N(u_1, \ldots, u_N, \alpha, t)]\| + \frac{1}{N}\sum_{i=1}^{N} |\tau \mathcal{R}_i(\alpha, t)| \leq 2C\tau^2(1+\alpha)^2 ,
\end{aligned} \tag{3.50}$$

which finished the proof after absorbing the coefficient 2 into C. □

We also need the usual regularity and stability assumptions to prove the accuracy of FLAVORs for (3.36).

Condition 3.4.3. *Assume that*

1. *$f_1, f_2, \ldots, f_N$ are Lipschitz continuous.*
2. *For all bounded initial condition $[u_1(0), \ldots, u_N(0)]$'s, the exact trajectories $([u_1(t), \ldots, u_N(t)])_{0 \leq t \leq T}$ (i.e., solution to (3.36)) are uniformly bounded in ϵ.*
3. *For all bounded initial condition $[u_1(0), \ldots, u_N(0)]$'s, the numerical trajectories $([\bar{u}_1(t), \ldots, \bar{u}_N(t)])_{0 \leq t \leq T}$ (defined by (3.43)) are uniformly bounded in ϵ, $0 < H \leq H_0$, $h \leq \min(H_0\epsilon, H)$.*

The following theorem shows the two-scale flow convergence of FLAVORs under the above conditions.

Theorem 3.4.1. *Consider FLAVOR trajectories in Definition 3.4.1. Under Conditions 3.4.1, 3.4.2 and 3.4.3, there exist $C > 0$, $\hat{C} > 0$ and $H_0 > 0$ independent from ϵ^{-1} and N, such that for $K/\hat{C} < H < H_0$, $h < H_0\epsilon$ and $t > 0$,*

$$\|x(t) - [\eta^t]^x(\bar{u}_1(t), \ldots, \bar{u}_N(t))\| \leq Ce^{Ct}\chi_1(u_1(0), \ldots, u_N(0), \epsilon, H, h) \tag{3.51}$$

and for all bounded and uniformly Lipschitz continuous test functions $\varphi : \mathbb{R}^N \mapsto \mathbb{R}$,

$$\left| \frac{1}{\Delta t} \int_t^{t+\Delta t} \varphi([\bar{u}_1(s), \ldots, \bar{u}_N(s)])\, ds - \int_{\mathbb{R}^p} \varphi([\eta^t]^{-1}(x(t), y))\mu^t(x(t), dy) \right| \\ \leq \chi_2(u_1(0), \ldots, u_N(0), \epsilon, H, h, \Delta t, t)(\|\varphi\|_{L^\infty} + \|\nabla\varphi\|_{L^\infty}) \tag{3.52}$$

where χ_1 and χ_2 are bounded functions converging towards zero as $\epsilon \leq H/(C \ln \frac{1}{H})$, $\frac{h}{\epsilon} \downarrow 0$, $\frac{\epsilon}{h}H \downarrow 0$ and $(\frac{h}{\epsilon})^2 \frac{1}{H} \downarrow 0$ (and $\Delta t \downarrow 0$ for χ_2); see Remark 2.1.5 for details about χ_1 and χ_2.

Recall notations: $NK = L$ is the fixed spatial width, $[\eta^t]^x$ and $[\eta^t]^{-1}$ respectively denote the x (slow) component and the inverse of the diffeomorphism η^t (defined in Condition 3.4.1), $x(t) = [\eta^t]^x(u_1(t), \ldots, u_N(t))$ corresponds to the slow component of the exact PDE solution sampled at grid points, and $u_i(t)$ and $\bar{u}_i(t)$ represent the exact and the FLAVOR approximation of the solution to the semi-discrete system with the remainders (3.36).

Proof. The proof of Theorem 3.4.1 is analogous to that of Theorems 2.1.1 and 2.1.2. The proof requires (3.46), which is guarantied from Condition 3.4.2 by Lemma 3.4.2. It is easy to check that the slow dependence on time of f, g, η and μ does not affect the proof given in Appendix A. □

Remark 3.4.3. *Condition 3.4.2 implies that the constant C in Theorem 3.4.1 does not depend on N or K. This is important because although using a finer*

mesh leads to a smaller K and a larger $N = L/K$, Condition 3.4.2 (which is equivalent to the accuracy of the semi-discrete approximation of the PDE) ensures that, as long as $K = \mathcal{O}(H)$ and $h \gg \epsilon H$, the constant C in the error bounds on the slow component (3.51) *and the fast component* (3.52) *will not blow up.*

Remark 3.4.4. *Observe that the application of the FLAVOR strategy does not require the identification of the diffeomorphism η (which may depend on the spatial discretization).*

3.5 On FLAVORizing characteristics

The convergence result of the previous section is based on the semi-discretization of the original PDE. PDEs and ODEs are also naturally connected via the method of characteristics, and therefore it is natural to wonder whether a numerical integration of those characteristics by FLAVORs would lead to an accurate approximation of the solution of the original PDE. The answer to this question will be illustrated by analyzing the following (generic) PDE:

$$\begin{cases} F(Du, u, q, \epsilon^{-1}) = 0, & q \in U \\ u(q) = \gamma(q), & q \in \Gamma \end{cases} \tag{3.53}$$

where $U \subset \mathbb{R}^d$ is the domain in which solution is defined, Γ and γ define initial/boundary conditions.

The following condition corresponds to assuming that characteristics are well-posed and the unknown has its value slowly changing along any characteristic.

Condition 3.5.1. *Assume that*

1. *The PDE $F(Du, u, q, \epsilon^{-1}) = 0$ admits characteristics:*

$$\dot{q} = f(q, z, \epsilon^{-1}) \tag{3.54}$$

$$\dot{z} = g(q, z) \tag{3.55}$$

$$u(q(t)) = z(t) \tag{3.56}$$

where $q \in U$ is a vector corresponding to coordinates of characteristics in the domain of the PDE, and z corresponds to the unknown's value along the characteristics.

2. *For arbitrary ϵ, any point in U is reachable from the initial condition via one and only one characteristic.*

The following conditions correspond to the assumption of existence of (possibly hidden) slow and locally ergodic fast variables for those characteristics.

Condition 3.5.2. *Consider ODE* (3.54)*. Assume that:*

1. *There exists a z-dependent diffeomorphism $\eta^z : q \mapsto [x, y]$ from $\mathbb{R}^d$ onto $\mathbb{R}^{d-p} \times \mathbb{R}^p$ with uniformly bounded C^1, C^2 derivatives with respect to both q and t, such that (x, y) satisfies (with $z(t)$ given by* (3.55)*)*

$$\begin{cases} \dot{x} &= f_1(x, y, z) \\ \dot{y} &= \epsilon^{-1} f_2(x, y, z) \end{cases} \tag{3.57}$$

where f_1, f_2, and g have bounded C^1 derivatives with respect to x, y and z, and $u([\eta^z]^{-1}(x, y))$ has bounded C^1 derivatives with respect to the (slow) variables x and z.

2. *There exists a family of probability measures $\mu^z(x, dy)$ on $\mathbb{R}^p$ indexed by $x \in \mathbb{R}^{d-p}$ and $z \in \mathbb{R}$, as well as a family of positive functions $T \mapsto E^z(T)$ satisfying $\lim_{T\to\infty} E^z(T) = 0$, such that for all x_0, y_0, z_0, T bounded and ϕ uniformly bounded and Lipschitz, the solution to*

$$\dot{Y}_t = f_2(x_0, Y_t, z_0) \qquad Y_0 = y_0 \tag{3.58}$$

satisfies

$$\Big|\frac{1}{T}\int_0^T \phi(Y_s)ds - \int_{\mathbb{R}^p} \phi(y)\mu^{z_0}(x_0, dy)\Big| \leq \chi^{z_0}\big(\|(x_0, y_0)\|\big) E^{z_0}(T)(\|\phi\|_{L^\infty} + \|\nabla\phi\|_{L^\infty}) \tag{3.59}$$

where $r \mapsto \chi^{z_0}(r)$ is bounded on compact sets, and μ^z has bounded derivative with respect to z in total variation norm.

The second item of Condition 3.5.2 corresponds to the assumption that the fast variable y is locally ergodic with respect to a family of measures μ drifted by the slow variables x and z.

The following lemma shows that, under the above conditions, the solution of PDE (3.53) is nearly constant on the orbit of the fast components (y) of any characteristic.

Lemma 3.5.1. *Under Conditions 3.5.1 and 3.5.2, for any fixed constant C_1 (independent of ϵ^{-1}), there exists a constant C_2 independent of ϵ^{-1}, such that for any $0 \leq t_1 \leq C_1$, $0 \leq t_2 \leq C_1$ and (fixed) x_0 and z_0,*

$$\left|u\left([\eta^{z_0}]^{-1}(x_0, Y(t_1))\right) - u\left([\eta^{z_0}]^{-1}(x_0, Y(t_2))\right)\right| \leq C_2\epsilon \tag{3.60}$$

where $Y(t_1)$ and $Y(t_2)$ are two points on the orbit of $\dot{Y}(t) = f_2(x_0, Y(t), z_0)$.

Proof. Under Conditions 3.5.1 and 3.5.2, it is known (we refer for instance to [239], or to Theorem 14, Section 3 of Chapter II of [258], or to [226]) that x and z converge as $\epsilon \to 0$ towards $\tilde{x}$ and $\tilde{z}$ defined as the solution to the following ODEs with initial condition x_0 and z_0

$$\begin{cases} \dot{\tilde{x}} &= \int f_1(\tilde{x}, y, \tilde{z})\mu^{\tilde{z}}(\tilde{x}, dy) \\ \dot{\tilde{z}} &= \int g([\eta^{\tilde{z}}]^{-1}(\tilde{x}, y), \tilde{z})\mu^{\tilde{z}}(\tilde{x}, dy) \end{cases} \tag{3.61}$$

Therefore, writing $y(t)$ the solution of $\dot{y} = \epsilon^{-1} f_2(\tilde{x}, y, \tilde{z})$, we have as $\epsilon \to 0$

$$u([\eta^{\tilde{z}(t)}]^{-1}(\tilde{x}(t), y(t))) \to \tilde{z}(t) \tag{3.62}$$

Now, taking the time derivative of $\hat{u} = u \circ \eta^{-1}$, we obtain

$$\hat{u}_x\dot{\tilde{x}} + \hat{u}_y\dot{y} + \hat{u}_z\dot{\tilde{z}} = \dot{\tilde{z}} + \dot{R}(\epsilon) \tag{3.63}$$

where $R(\epsilon)$ is a function of t that goes to 0 as $\epsilon \to 0$.

Furthermore,

$$\begin{aligned}\dot{Y}(t) &= f_2(x_0, Y(t), z_0) \\ &= f_2(\tilde{x}(\epsilon t), y(\epsilon t), \tilde{z}(\epsilon t)) + \frac{\partial f_2}{\partial \tilde{x}}(\tilde{x}(\epsilon t) - x_0) + \frac{\partial f_2}{\partial \tilde{z}}(\tilde{z}(\epsilon t) - z_0) + \frac{\partial f_2}{\partial y}(y(\epsilon t) - Y(t)) \\ &\quad + o(\epsilon) + o(y(\epsilon t) - Y(t))\end{aligned}$$

By Taylor expansion, $\tilde{x}(\epsilon t) - x_0$ and $\tilde{z}(\epsilon t) - z_0$ are obviously $\mathcal{O}(\epsilon)$. Applying Gronwall's lemma, we also obtain that $y(\epsilon t) - Y(t) = \mathcal{O}(\epsilon)$. Therefore,

$$\dot{Y}(t) = f_2(\tilde{x}(\epsilon t), y(\epsilon t), \tilde{z}(\epsilon t)) + \mathcal{O}(\epsilon) = \epsilon \dot{y}(t) + o(\epsilon) \tag{3.64}$$

Combining (3.63) with (3.64), we obtain

$$\begin{aligned}u\left(\eta^{-1}(x_0, Y(t_1))\right) - u\left(\eta^{-1}(x_0, Y(t_2))\right) &= \int_{t_1}^{t_2} \hat{u}_y \cdot \dot{Y}(t)\, dt = \epsilon \int_{t_1}^{t_2} \hat{u}_y \cdot \dot{y}\, dt + o(\epsilon) \\ &= \epsilon \left(\int_{t_1}^{t_2} (\dot{\tilde{z}} - \hat{u}_x \dot{\tilde{x}} - \hat{u}_z \dot{\tilde{z}})\, dt + R(\epsilon)\Big|_{t_1}^{t_2} \right) + o(\epsilon)\end{aligned} \tag{3.65}$$

Since $\hat{u}_x$, $\dot{\tilde{x}}$, $\hat{u}_t$ and $\dot{\tilde{z}}$ are bounded, and $R(\epsilon)$ is vanishing (and hence bounded), we conclude that the right hand side is $\mathcal{O}(\epsilon)$. □

Condition 3.5.3. *Assume that the domain U is bounded (independently from ϵ^{-1}).*

Lemma 3.5.2. *If Conditions 3.5.1, 3.5.2, and 3.5.3 hold, then every point in U is reachable by a characteristic from the initial condition in bounded time (independently from ϵ^{-1}).*

Proof. From Condition 3.5.1, we already know that every point is reachable, and therefore it suffices to show that hitting times do not blow up as $\epsilon \to 0$. Since $x(\cdot)$ converges to $\tilde{x}(\cdot)$ (see proof of Lemma 3.5.1), by considering the x component of the characteristic (projected by η), it becomes trivial to show that the hitting time converges to a fixed value (and hence, does not blow up). Using Condition 3.5.3,

we conclude that that any point in U can be hit in (uniformly) bounded time from the initial condition. □

Analogously to the Integrator 3.4.1, a legacy integrator for (3.54) and (3.55) can be FLAVORized, and shown to be convergent under regularity and stability conditions (analogous to Condition 3.4.3) requiring f_1, f_2 and g to be Lipschitz continuous and $\tilde{q}(t)$ and $\tilde{z}(t)$ to be bounded. The convergence result is analogous to Theorem 3.4.1, modulo the following change of notation: the slow index is now z instead of t, the original coordinates are q instead of u_i, the vector field of the original coordinates is f instead of f_i, and the dynamics of the slow index comes from the nontrivial drift of $\dot{z} = g(q, z)$ instead of the trivial $\dot{t} = 1$. We define $\tilde{u}(\tilde{q}(t)) := \tilde{z}(t)$ for all t on each FLAVORized characteristic $[\tilde{q}(t), \tilde{z}(t)]$. Naturally, $\tilde{u}$ is only defined at discrete points in the domain U. These discrete points, however, densely 'fill' the space in the sense that (as shown by the proof of the following theorem) FLAVORied characteristics remain very close to exact characteristics (x components are close in Euclidean distance, and y components are close as well in terms of orbital distance induced by the infimum of point-wise Euclidean distances).

By the two-scale convergence theorem, we can quantify for each characteristic: the strong convergence of its slow coordinate and the unknown's value along it, and the weak convergence of its fast coordinate. Finally, the ODE FLAVORization error bounds of each characteristic collectively transfers to error bounds of the PDE approximation by considering the entire family of characteristics starting from all points (in the initial condition).

Theorem 3.5.1. *Write $\tilde{u}(\tilde{q})$ the solution obtained by FLAVORizing all characteristics. Under Conditions 3.5.1, 3.5.2, 3.5.3, a consistency condition analogous to Condition 3.4.2, and a regularity and stability condition analogous to Condition 3.4.3 (under a change of notation as described above), there existsW a constant C independent of ϵ^{-1} and $q_0 \in \Gamma$, such that*

$$|\tilde{u}(\tilde{q}) - u(\tilde{q})| \leq C\chi_1(q_0, \gamma(q_0), \epsilon, \delta, \tau)(1 + \chi_2(q_0, \gamma(q_0), \epsilon, \delta, \tau, T, t)) \tag{3.66}$$

for any $\tilde{q}$ on any FLAVORized characteristic, where $q_0 \in \Gamma$ and $\gamma(q_0)$ correspond to the initial condition that leads to $\tilde{q}$ via a FLAVORized characteristic, and χ_1 and χ_2 are vanishing error bound functions.

Remark 3.5.1. *When Γ is compact (such as in the case of periodic boundary condition), χ_1 and χ_2 can be further chosen to be independent of q_0 (hence $\tilde{q}$) by taking a supremum over Γ.*

Proof. By Condition 3.5.1, each $q \in U$ can be traced back to $q_0 \in \Gamma$ through a characteristic. By Lemma 3.5.2, the characteristic starting from q_0 reaches q in bounded time T. Using the two-scale convergence of the FLAVORization of these characteristics (a result analogous to Theorem 3.4.1), we deduce that the approximation error associated with $\tilde{z}_T$ (on each FLAVORized characteristic) can be bounded by $C\chi_1$ (with respect to the true value $u(q) = z_T$, the error Ce^{CT} has been replaced by C because T is bounded).

Now observe that $\tilde{q}_T \neq q_T$, where $\tilde{q}_T$ is the coordinate of the FLAVORized characteristics starting from q_0. As before, let $[x_T, y_T] = \eta(q_T)$ and $[\tilde{x}_T, \tilde{y}_T] = \eta(\tilde{q}_T)$. The error on the slow component is $\|x_T - \tilde{x}_T\| \leq C\chi_1$. The possibly large error on the fast component is not a problem because we can look for a near-by point on the fast orbit with introducing only an $\mathcal{O}(\epsilon)$ error on the unknown's value (Lemma 3.5.1):

$$\begin{cases} u(\eta(x_T, y_T)) = u(\eta(x_T, y_T^*)) + \mathcal{O}(\epsilon) \\ y_T^* = \arg\min_{Y_t | \dot{Y}_t = f(x_T, Y_t)} \|\tilde{y}_T - Y_t\| \end{cases} \tag{3.67}$$

Since $\|\tilde{x}_T - x_t\|$ is small, the local ergodic measures that represent the orbits given by $\dot{Y}_t = f(x_T, Y_t)$ and $\dot{Y}_t = f(\tilde{x}_T, Y_t)$ will be small: $\|\mu(x_T, dy) - \mu(\tilde{x}_T, dy)\|_{\text{T.V.}} \leq C\chi_1\chi_2$ is by chain rule. Because $\tilde{y}_T$ is on the orbit of $\dot{Y}_t = f(\tilde{x}_T, Y_t)$, we will have $\|y_T^* - \eta^y(\tilde{q}_T)\| \leq C\chi_1\chi_2$.

All together, we obtain

$$\begin{aligned}
|\tilde{u}(\tilde{q}_T) - u(\tilde{q}_T)| &= |\tilde{z}_T - u(\tilde{q}_T)| \\
&\leq |\tilde{z}_T - u(q)| + |u(q_T) - u(\tilde{q}_T)| \\
&\leq C\chi_1 + C\|\nabla(u \circ \eta)\|_\infty \left(\|x_T - \eta^x(\tilde{q}_T)\| + \|y_T - \eta^y(\tilde{q}_T)\|\right) \\
&\leq C\chi_1 + C(\chi_1 + \chi_1\chi_2) = C\chi_1 + C\chi_1\chi_2 \qquad (3.68)
\end{aligned}$$

□

Remark 3.5.2. *To keep the presentation concise, we have written C to indicate all constants that do not depend on essential parameters.*

Remark 3.5.3. *As shown above, u will be captured strongly. Du, on the other hand, depends on a derivative with respect to the fast variable, and therefore will only be convergent in a weak sense.*

Relevance to an error analysis for PDE-FLAVORs: The above result guarantees the convergence of FLAVORized characteristics. It is also possible to establish an error bound on the difference between a specific PDE-FLAVOR discretization and the approximation given by the above FLAVORized characteristics (and hence prove the convergence of this specific PDE-FLAVOR discretization). Such an error bound could be obtained by first transforming FLAVORized characteristics to PDE-FLAVOR grid points via interpolating functions, and then using the fact that coordinate transformations do not affect the efficiency of FLAVORs. Many details are left to be filled in.

Chapter 4

Quadratic and quasi-quadratic stiff potentials

Significant research has been done for coarse-timing-stepping of mechanical systems with quadratic stiff potentials (see Section 1.3.3), i.e., with Hamiltonian written as

$$\mathcal{H}(q,p) = \frac{1}{2}p^T M^{-1} p + V(q) + \frac{\epsilon^{-1}}{2} q^T K q, \tag{4.1}$$

where M and K are constant positive-definite symmetric matrices and $V(\cdot)$ is an arbitrary function. This type of system is also frequently referred to as highly-oscillatory second-order differential equations.

Highly-oscillatory mechanical systems call for a specialized treatment, even though general multiscale strategies such as FLAVORization (Chapters 2 and 3) apply. On the one hand, computations could be further accelerated in this special case, because it is possible to use a macroscopic timestep for symplectic integrations — although both are independent of the stiffness in the equation, a macroscopic step is an order of magnitude larger than a mesoscopic step, which is required by general multiscale methods based on averaging, such as FLAVORs, HMM and equation-free (notice only FLAVORs are symplectic).

On the other hand, all position variables, including the fast ones, could be captured numerically in a strong sense. This seems implausible at a first glance, because a single integration step (macroscopic) spans over many periods of the fast process. However, this peculiarity, as well as the previously mentioned acceleration by the meso-to-macro up-scaling, could be simultaneously explained: in FLAVORs

we sample the fast process in order to get its effective contribution, but for highly oscillatory mechanical systems it is known that the fast process approximates harmonic oscillations at each step. Because of this, both the weak convergence of FLAVORs on fast variables and the strong convergence of methods in the chapter are natural. In fact, the weak convergence of FLAVORs is already optimal, because in general, a signal of B hertz needs $2B$ points for its representation due to Nyquist-Shannon sampling theorem [251], and these $2B$ points are clearly more than what a mesoscopic step could provide. A more careful inspection shows, however, that this theorem is only a sufficient but not necessary condition. If additional a priori information is available, such as a low dimensional set of bases for the signal, then the signal can be reconstructed by fewer parameters[1]. This is indeed why the fast process in highly oscillatory mechanical systems could be captured by macroscopic steps, because in this case bases are known to be harmonic functions. For general unknown fast processes, which are what FLAVORs aim at, however, a low dimensional set of bases does not exist a priori.

Regarding integrations of this type, we will present three of our contributions:

1. The impulse method [297, 124, 286] is one of the prevailing methods that enables such an integration. It has also been 'mollified', i.e., a filter on the slow force is introduced, so that the resulting method has better stability and accuracy.

   We try to find a variational principle for the impulse method, but end up rediscovering one of its mollified versions. This possibly could facilitate more analysis on the mollified impulse method, including backward error analysis and a study of momentum maps, but the generality of this new variational principle is much beyond the variational derivation of a mollified impulse method: it should work for imposing arbitrary assumptions on the form of the solutions (for the case of impulse, the assumption is that each half-step update corresponds to the flow of harmonic oscillators). Details of

[1] We refer to compressive sensing [58, 83] for a discussion.

this principle will be given in Section 4.1, but its applications remain to be investigated.

2. There have been, of course, various error analyses on impulse methods, for instance [109, 123, 240], but either (i) position and momentum are shown to be convergent at different orders, or (ii) a filter on the slow force is needed so that position and momentum can be considered together in the analysis, in which case, however, the method is no longer the original impulse but the mollified ones.

   We use a different perspective to view the original impulse method as a splitting method, and by doing this we are able to [276] (i) propose a new error analysis that treats position and momentum simultaneously; (ii) generalize the impulse method to the general Langevin case with slow force, fast (linear) force, (fast) friction, and noise, all present at the same time (the impulse method has been extended to Langevin dynamics [255], but the method there only considers a slow force and could not handle fast frictions); (iii) show for the first time that the original impulse method is not only symplectic but also variational.

3. Lastly, but perhaps most significantly (it is stated last simply because it needs the previous results), we use the perspective of splitting and the error analysis introduced above to propose coarse-step-integrators of mechanical systems with quasi-quadratic stiff potentials, i.e., with Hamiltonian

$$\mathcal{H}(q,p) = \frac{1}{2}p^T M^{-1} p + V(q) + \frac{\epsilon^{-1}}{2}[q^{fast}]^T K(q^{slow}) q^{fast}, \tag{4.2}$$

   We also propose efficient and symplectic matrix exponentiation algorithms to enable such integrations, but these exponentiation algorithms are generic, and their applications are not limited to numerical integrations.

   As a result, (possibly high-dimensional) mechanical systems with quasi-quadratic stiff potentials can be accurately and symplectically integrated

using macroscopic timesteps. Our approach is so far the only method that could do so (see [82] for a method for systems that have either one varying fast frequency or several constant frequencies).

4.1 Hamilton-Pontryagin-Marsden principle

4.1.1 The general variational principle

Hamilton-Pontryagin principle is a variational principle that generalizes the well-known Hamilton principle of 'least' action. Specifically, given the Lagrangian $\mathcal{L} : Q \times TQ \mapsto \mathbb{R}$ of a mechanical system in a configuration space Q, the equation of motion will be given by the critical point $(q(t), v(t), p(t))$ of the following functional:

$$\mathcal{S} = \int_0^T \mathcal{L}(q, v) + \langle p, \dot{q} - v \rangle \, dt \tag{4.3}$$

where naturally $q(t) \in Q$, $v(t) \in TQ$, and $p(t) \in T^*Q$. It is not difficult to see that they indeed correspond to position, velocity and momentum.

Variational integrators could be derived by discretizing the action using quadratures, and then taking the variation with respect to finite many arguments. For instance, one simplest discretization (based on a left-point rule) would be:

$$\mathcal{S}_d = \sum_{i=0}^{N-1} h \left(\mathcal{L}(q_i, v_i) + \left\langle p_i, \frac{q_{i+1} - q_i}{h} - v_i \right\rangle \right), \tag{4.4}$$

where h is the timestep satisfying $Nh = T$.

Now suppose there is some a priori knowledge about the form of the solution, i.e., $q_{i+1} = f(q_i, v_i)$ for some function f, which we wish to enforce and incorporate into the variational principle. It can be seen that (4.4) in fact uses an assumption that $q_{i+1} = q_i + hv_i$. In the general case, we will simply replace that by f, and obtain the following discrete action:

$$\mathcal{S}_d = \sum_{i=0}^{n-1} h \left(\mathcal{L}(q_i, v_i) + \left\langle p_i, \frac{q_{i+1} - f(q_i, v_i)}{h} \right\rangle \right), \tag{4.5}$$

I wish to call the discrete action (4.5) accompanied by the usual variational principle of $\delta\mathcal{S}_d = 0$ the **Hamilton-Pontryagin-Marsden principle**, in order to acknowledge Professor Jerrold Eldon Marsden's immortal contributions to continuous and discrete mechanics.

4.1.2 An example application to quadratic stiff potentials

An easiest example of the applications of Hamilton-Pontryagin-Marsden principle (its action defined in (4.5)) is that it gives the exact solution to a harmonic oscillator, although the principle itself is only a numerical approximation.

Specifically, if the Lagrangian is $\mathcal{L}(q,v) = \frac{1}{2}v^2 + \frac{1}{2}\omega^2 q^2$, and an a priori knowledge of $q_{i+1} = \cos(\omega h)q_i + \sin(\omega h)/\omega v_i$ is available, by solving

$$\begin{cases} \partial_{q_i}\mathcal{S}_d = 0 \\ \partial_{v_i}\mathcal{S}_d = 0 \\ \partial_{p_i}\mathcal{S}_d = 0 \end{cases} \tag{4.6}$$

for all i and eliminating p_i, we obtain:

$$\begin{cases} q_{i+1} = \cos(\omega h)q_i + \frac{\sin(\omega h)}{\omega}v_i \\ v_{i+1} = -\omega\sin(\omega h)q_i + \cos(\omega h)v_i \end{cases}, \tag{4.7}$$

which is the exact solution of the system (notice that we get the correct velocity for free).

Now, if the Lagrangian is $\mathcal{L}(q,v) = \frac{1}{2}v^2 + \frac{1}{2}\omega^2 q^2 + V(q)$, i.e., the one considered by impulse and mollified impulse methods, and the same a priori knowledge is again used, then by applying Hamilton-Pontryagin-Marsden principle (its action defined in (4.5)) we obtain:

$$\begin{cases} q_{i+1} = \cos(\omega h)q_i + \frac{\sin(\omega h)}{\omega}v_i \\ v_{i+1} = -\omega\sin(\omega h)q_i + \cos(\omega h)v_i - \frac{\sin(\omega h)}{\omega}\nabla V(q_{i+1}) \end{cases} \tag{4.8}$$

Notice that a first order version of the impulse method is (see Section 4.2):

$$\begin{cases} q_{i+1} = \cos(\omega h) q_i + \frac{\sin(\omega h)}{\omega} v_i \\ v_{i+1} = -\omega \sin(\omega h) q_i + \cos(\omega h) v_i - h \nabla V(q_{i+1}) \end{cases}, \tag{4.9}$$

which is apparently the limit of (4.8) as $\omega \to 0$.

Curiously enough, (4.8) coincides with a first-order version of the mollified impulse method that corresponds to the LongAverage filter on the slow force (see [109]). Although we will also show that impulse method is variational (Section 4.2), (4.9) is interesting in the sense that its variational formulation is in a close form (as opposed to an infinite series expansion involving Poisson brackets for the case of the original impulse method). This opens the possibilities for various analyses, such as conservation of momentum maps, modulated Fourier expansion, and backward error analysis.

4.2 Stochastic impulse methods and error analysis in energy norm

Most results in this section can be found in a submitted manuscript [276].

4.2.1 Methodology

We will directly consider the Langevin setting, for the Hamiltonian setting is a degenerate case of it. Namely, consider numerical integrations of the following stiff Langevin SDEs

$$\begin{cases} M dq &= p dt \\ dp &= -\nabla V(q) dt - \epsilon^{-1} K q dt - c p dt + \sigma dW \end{cases} \tag{4.10}$$

where $0 < \epsilon \ll 1$, $q \in \mathbb{R}^d$, $p \in \mathbb{R}^d$, K is a positive definite $d \times d$ matrix, c and σ are positive semi-definite $d \times d$ matrices, respectively indicating viscous damping coefficients and amplitudes of noises. We restrict ourselves to Euclidean phase

spaces, although the method is readily generalizable to manifolds. In addition, we require that matrices K and c commute; a special case satisfying this requirement is c being a scalar.

In the case of no noise no friction ($c = 0$ and $\sigma = 0$), the system degenerates to a deterministic mechanical system with Hamiltonian (4.1).

Also, the method as well as its properties (e.g., uniform convergence) generalizes to open systems:

$$\begin{cases} Mdq &= pdt \\ dp &= F(q)dt - \epsilon^{-1}Kqdt - cpdt + \sigma dW \end{cases} \tag{4.11}$$

but we stick to (4.10) for simplicity in descriptions.

Denote by $\phi^f(\tau) : \big(q^f(t), p^f(t)\big) \mapsto \big(q^f(t+\tau), p^f(t+\tau)\big)$ and $\phi^s(\tau) : (q^s(t), p^s(t)) \mapsto (q^s(t+\tau), p^s(t+\tau))$ respectively the τ-flow maps of the autonomous SDE systems

$$\begin{cases} Mdq^f &= p^f dt \\ dp^f &= -\epsilon^{-1}Kq^f dt - cp^f dt + \sigma dW \end{cases} \tag{4.12}$$

and

$$\begin{cases} Mdq^s &= 0 \\ dp^s &= -\nabla V(q^s)dt \end{cases} \tag{4.13}$$

Since the first system is a linear SDE and the second is a free drift, flows of both can be obtained exactly.

Then Stochastic Impulse Methods (SIMs) are defined via compositions of ϕ^f and ϕ^s. Here are several examples of SIMs with a timestep H:

Integrator 4.2.1. ***1st-order SIM in the*** $c = 0$, $\sigma = 0$ ***case****, is given by the*

one step update of $\phi^s(H) \circ \phi^f(H)$*:*

$$\begin{cases} q_{k'} &= A_{11}(H)q_k + A_{12}(H)p_k \\ p_{k'} &= A_{21}(H)q_k + A_{22}(H)p_k \\ q_{k+1} &= q_{k'} \\ p_{k+1} &= p_{k'} - H\nabla V(q_{k'}) \end{cases}$$

$$\textit{where} \begin{bmatrix} A_{11}(H) & A_{12}(H) \\ A_{21}(H) & A_{22}(H) \end{bmatrix} = \exp \begin{bmatrix} 0 & M^{-1}H \\ -\epsilon^{-1}KH & 0 \end{bmatrix},$$

$$\begin{cases} q_0 = q(0) \\ p_0 = p(0) \end{cases}$$

Remark 4.2.1. *The other 1st-order SIM, as the above's dual, can be obtained via the one step update* $\phi^f(H) \circ \phi^s(H)$*. Both these 1st-order composition schemes are well known as the Lie-Trotter splitting [285].*

Integrator 4.2.2. ***1st-order SIM in the full Langevin case****, given by the*

same one step update $\phi^s(H) \circ \phi^f(H)$*:*

$$\begin{cases} q_{k'} &= B_{11}(H)q_k + B_{12}(H)p_k + Rq_k(H) \\ p_{k'} &= B_{21}(H)q_k + B_{22}(H)p_k + Rp_k(H) \\ q_{k+1} &= q_{k'} \\ p_{k+1} &= p_{k'} - H\nabla V(q_{k'}) \\ \begin{bmatrix} Rq_k(H) \\ Rp_k(H) \end{bmatrix} &\sim \mathcal{N}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma^2_{11}(H)) & \Sigma^2_{12}(H)) \\ \Sigma^2_{21}(H)) & \Sigma^2_{22}(H)) \end{bmatrix}), \textit{i.i.d.} \end{cases}$$

$$\textit{where} \begin{bmatrix} B_{11}(H) & B_{12}(H) \\ B_{21}(H) & B_{22}(H) \end{bmatrix} = \exp \begin{bmatrix} 0 & M^{-1}H \\ -\epsilon^{-1}KH & -cH \end{bmatrix},$$

$$\begin{cases} q_0 = q(0) \\ p_0 = p(0) \end{cases},$$

$$\begin{cases} \Sigma^2_{11}(H) = \int_{s=0}^{H} \left(B_{12}(H-s)\sigma\sigma^T B^T_{12}(H-s)\right) ds \\ \Sigma^2_{12}(H) = \int_{s=0}^{H} \left(B_{12}(H-s)\sigma\sigma^T B^T_{22}(H-s)\right) ds \\ \Sigma^2_{21}(H) = \int_{s=0}^{H} \left(B_{22}(H-s)\sigma\sigma^T B^T_{12}(H-s)\right) ds \\ \Sigma^2_{22}(H) = \int_{s=0}^{H} \left(B_{22}(H-s)\sigma\sigma^T B^T_{22}(H-s)\right) ds \end{cases}$$

Remark 4.2.2. $\begin{bmatrix} Rq_k(H) \\ Rp_k(H) \end{bmatrix}$ *indicates the value of* $\int_{s=0}^{H} B(H-s) \begin{bmatrix} 0 \\ \sigma dW_s \end{bmatrix}$ *and hence is a vectorial normal random variable with zero mean and covariance of* $\begin{bmatrix} \Sigma^2_{11}(H) & \Sigma^2_{12}(H) \\ \Sigma^2_{21}(H) & \Sigma^2_{22}(H) \end{bmatrix}$.

Integrator 4.2.3. ***2nd-order SIM in the full Langevin case****, given by the*

one step update $\phi^s(H/2) \circ \phi^f(H) \circ \phi^s(H/2)$*:*

$$\begin{cases} q_{k'} &= q_k \\ p_{k'} &= p_k - \frac{H}{2}\nabla V(q_k) \\ q_{k''} &= B_{11}(H)q_{k'} + B_{12}(H)p_{k'} + Rq_k(H) \\ p_{k''} &= B_{21}(H)q_{k'} + B_{22}(H)p_{k'} + Rp_k(H) \\ q_{k+1} &= q_{k''} \\ p_{k+1} &= p_{k''} - \frac{H}{2}\nabla V(q_{k''}) \end{cases}$$

Remark 4.2.3. *This uses the 2nd-order composition scheme known as the Strang or Marchuk splitting [266, 191]. When no noise or friction, i.e.* $c = 0$ *and* $\Sigma = 0$*, the resulting integrator degenerates to the prevailing Verlet-I/r-RESPA impulse method [124, 286].*

Remark 4.2.4. *Higher order SIMs can be obtained systematically since generic way for constructing higher order splitting/composition schemes exists (see for example [128]). For instance, a 4th order SIM is given by* $\phi^s(cH/2) \circ \phi^f(cH) \circ \phi^s((1-c)H/2) \circ \phi^f((1-2c)H) \circ \phi^s((1-c)H/2) \circ \phi^f(cH) \circ \phi^s(cH/2)$ *where* $c = \frac{1}{2-2^{1/3}}$ *[213].*

4.2.2 Preserved structures

In the case of $c = 0$ and $\sigma = 0$, since ϕ^s and ϕ^f are the exact flows of Hamiltonian systems, they are symplectic. Therefore, SIMs, as compositions of the two, are symplectic.

Moreover, SIMs are not only symplectic but variational, in the sense that their equations of motion are obtained as the critical point of a well-defined action, which is the integral of a discrete Lagrangian and dependent on finitely many discrete degrees of freedom. In fact, SIMs exactly preserve a Hamiltonian, which could be obtained from $\mathcal{H}_1(q,p) := V(q)$ and $\mathcal{H}_2 := \frac{1}{2}p^T M^{-1} p + \frac{\epsilon^{-1}}{2} q^T K q$ via Poisson brackets, because a SIM is a composition of two flows that correspond to the split Hamiltonian systems (see [128] for how Hamiltonian splitting results in a

new Hamiltonian). If a Lagrangian aspect is needed, a non-degenerate Legendre transformation will do the job. Nevertheless, neither the corresponding Hamiltonian or the Lagrangian will be in a closed form, because they are both expressed in forms of infinite series [128].

When noise and friction are present, SIMs are quasi-symplectic for RL1 and RL2 in [205] can be easily checked to be true, i.e., they degenerate to symplectic methods if friction is set equal to zero and the Jacobian of the flow map is independent of (q, p). In addition, if c is isotropic, then SIMs are conformally symplectic, i.e., they preserve the precise symplectic area change associated to the flow of inertial Langevin processes [202]. These properties are consistent with the numerically observed convergence (in distribution) towards the Boltzmann-Gibbs invariant measure.

4.2.3 Uniform convergence

In the case of $c = 0$ and $\sigma = 0$, convergence of SIMs is guaranteed by the general construction of splitting schemes. In the full Langevin setting, analogous convergence results for the same splitting schemes can be easily obtained using generators of SDEs. By this approach, however, the error bound will contain the scaling factor ϵ^{-1} and therefore restrain the timestep from being large.

Instead, we seek for uniform convergence results, i.e., error bounds that don't depend on ω. It turns out a uniform error bound on both position q and momentum p holds, but in a special norm called scaled energy norm, which translates to a uniform error bound on q in Euclidean norm but a non-uniform bound on p in Euclidean norm.

Definition 4.2.1. ***Scaled energy norm:***

$$\Omega \triangleq \epsilon^{-1/2}\sqrt{K}$$

$$\| \begin{bmatrix} q \\ p \end{bmatrix} \|_E \triangleq \| \begin{bmatrix} q \\ \Omega^{-1}p \end{bmatrix} \|_2 = \sqrt{q^T q + \epsilon p^T K^{-1} p}$$

This is well defined because K is positive definite.

Condition 4.2.1. *We will prove a uniform bound on the scaled energy norm of the global error of Integrator 4.2.2 if the following conditions hold:*

1. *Matrices c and K commute. A special case could be c being a scalar.*
2. $\lim_{\epsilon\to 0}\sqrt{\epsilon}\|c\|_2 \le C$ *for some constant C independent of ϵ, i.e.* $c \le \mathcal{O}(\epsilon^{-1/2})$.
3. σ *is independent of ϵ^{-1}, in the sense that* $\lim_{\epsilon\to 0}\epsilon^p\|\sigma\|_2 = 0$ *for any* $p > 0$.
4. *In the integration domain of interest $\nabla V(\cdot)$ is bounded and Lipschitz continuous with coefficient L, i.e.* $\|\nabla V(a) - \nabla V(b)\|_2 \le L\|a-b\|_2$.
5. *Denote by $x(T) = (q(T), p(T))$ the exact solution to (4.10), and $x_T = (q_T, p_T)$ the discrete numerical trajectory given by Integrator 4.2.2, then $\mathbb{E}\|x(T)\|_2^2 \le C$ and $\mathbb{E}\|x_T\|_2^2 \le C$ for some constant C independent of ϵ^{-1} but dependent on initial condition $\mathbb{E}\|\begin{bmatrix} q_0 \\ p_0 \end{bmatrix}\|_2^2$, amplitude of noise σ and friction c. (This is the traditional stability requirement, but it could also be understood as a bounded energy requirement.)*

Remark 4.2.5. *Notice that the damping coefficient c is allowed to be large (stiff). The general GLA [41] approach of constructing a Langevin integrator from a symplectic scheme by composing an Ornstein-Uhlenbeck flow with the symplectic integrator will not allow a macroscopic timestep in this case of fast dissipation, but SIMs do not have such a problem.*

Theorem 4.2.1. *If Condition 4.2.1 holds, the 1^{st} order SIM (Integrator 4.2.2) for multiscale Langevin system (4.10) ($c \neq 0$, $\sigma \neq 0$) has in mean square sense a uniform global error of $\mathcal{O}(H^{1/2})$ in q and a non-uniform global error of $\epsilon^{-1/2}\mathcal{O}(H^{1/2})$ in p, given a fixed total simulation time $T = NH$:*

$$(\mathbb{E}\|q(T) - q_T\|_2^2)^{1/2} \le CH^{1/2} \tag{4.14}$$

$$(\mathbb{E}\|p(T) - p_T\|_2^2)^{1/2} \le \epsilon^{-1/2}\|\sqrt{K}\|_2 CH^{1/2} \tag{4.15}$$

where $q(T), p(T)$ is the exact solution and q_T, p_T is the numerical solution; C is a positive constant independent of ϵ^{-1} but dependent on simulation time T, scaleless elasticity matrix K, scaled damping coefficient $\sqrt{\epsilon}c$ ($\mathcal{O}(1)$), amplitude of noise σ, slow potential energy $V(\cdot)$, and initial condition $\mathbb{E}\| \begin{bmatrix} q_0 \\ p_0 \end{bmatrix} \|_2^2$.

Proof. We refer to Appendix A.3. □

Remark 4.2.6. *By looking at the proof, one can be assured that all convergence results of SIMs apply to situations where the deterministic system is in a more general form of $M\frac{d^2}{dt^2}q = -\epsilon^{-1}Kq + F(q)$, where $F(q)$ doesn't have to be $-\nabla V(q)$.*

In the special case of Hamiltonian system, the same integrator gains 1/2 more order of accuracy.

Condition 4.2.2. *We will prove a uniform bound on the scaled energy norm of the global error of Integrator 4.2.1 if the following conditions hold:*

1. *In the integration domain of interest $\nabla V(\cdot)$ is bounded and Lipschitz continuous with coefficient L, i.e. $\|\nabla V(a) - \nabla V(b)\|_2 \leq L\|a - b\|_2$.*
2. *Denote by $x(T) = (q(T), p(T))$ the exact solution to (4.10) with $c = 0$ and $\sigma = 0$, and $x_T = (q_T, p_T)$ the discrete numerical trajectory given by Integrator 4.2.1, then $\|x(T)\|_2^2 \leq C$ and $\|x_T\|_2^2 \leq C$ for some constant C independent of ϵ^{-1} but dependent on initial condition $\| \begin{bmatrix} q_0 \\ p_0 \end{bmatrix} \|_2^2$. (This is the traditional stability requirement, but it could also be understood as a bounded energy requirement.)*

Theorem 4.2.2. *If Condition 4.2.2 holds, the 1^{st} order SIM (Integrator 4.2.1) for multiscale Hamiltonian system ((4.10) with $c = 0$, $\sigma = 0$) has a uniform global error of $\mathcal{O}(H)$ in q and a non-uniform global error of $\epsilon^{-1/2}\mathcal{O}(H)$ in p, given a fixed total simulation time $T = NH$:*

$$\|q(T) - q_T\|_2 \leq CH \tag{4.16}$$

$$\|p(T) - p_T\|_2 \leq \epsilon^{-1/2}\|\sqrt{K}\|_2 CH \tag{4.17}$$

where $q(T), p(T)$ is the exact solution and q_T, p_T is the numerical solution; C is a positive constant independent of ϵ^{-1} but dependent on simulation time T, scaleless elasticity matrix K, slow potential energy $V(\cdot)$ and initial condition $\| \begin{bmatrix} q_0 \\ p_0 \end{bmatrix} \|_2$.

Proof. It follows by simplifying the proof of Theorem 4.2.1. □

Remark 4.2.7. *These results are to our knowledge the first error analysis of the impulse method that unites both position and momentum without the introduction of a slow force filter (i.e., mollification).*

4.2.4 Stability

As one sees from Condition 4.2.1 and 4.2.2 (as another nonlinear demonstration of Lax equivalence theorem [174]), stability is necessary for global convergence. Instability could either come from the problem itself (not all SDEs have bounded solutions in the mean square sense), or from imperfection in numerical integration schemes. Here consider the latter possibility only. It is shown that impulse methods are not unconditionally stable [109], and its improvement, mollified impulse methods, are still susceptible to instability intervals (although narrower) in a linear example [55]. Nevertheless, instability intervals of impulse method are already narrow regions; for instance, the first instability interval in the stiff example considered by [55] is $0.544 < H < 0.553$. It is intuitive that instability intervals for the stochastic case with damping or higher-order schemes will not be wider. Therefore, one could still choose a large timestep H in SIMs without hitting the instability, by at most a few integration tryouts with slightly varied H values.

Remark 4.2.8. *Stochastic impulse method could be mollified by using Hamilton-Pontryagin-Marsden principle (Section 4.1) if additional stability is desired. This is one possible future direction.*

4.2.5 A stochastic numerical example

Consider a "Wall – linear stiff Spring – Mass – nonlinear soft Spring – Mass" system with both masses under isotropic noise and friction (Figure 4.1). The Hamiltonian is $H(x, y, p_x, p_y) = \frac{1}{2}p_x^2 + \frac{1}{2}p_y^2 + \frac{1}{2}\omega^2 x^2 + \frac{1}{4}(y-x)^4$ and the governing equations write as:

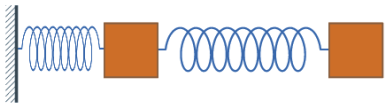
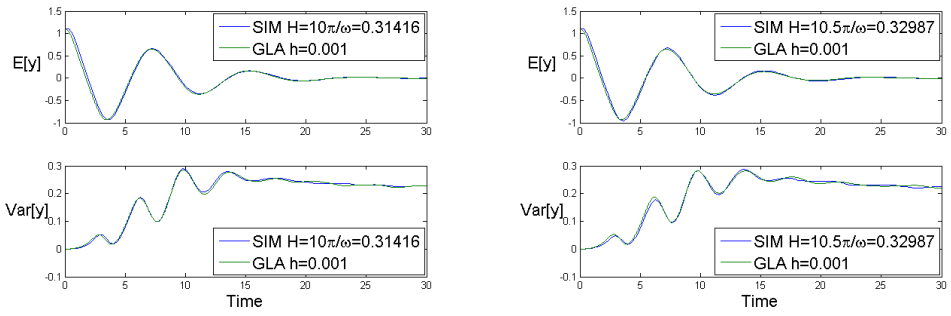


Figure 4.1: 2-spring systems

$$\begin{cases} dx &= p_x dt \\ dy &= p_y dt \\ dp_x &= -(\omega^2 x + (x-y)^3)dt - cp_x dt + \sigma dW_t^1 \\ dp_y &= -(y-x)^3 dt - cp_y dt + \sigma dW_t^2 \end{cases}$$



(a) ***Full period case:*** $\sin(\omega H) = 0$ **(b)** ***Quarter period case:*** $\cos(\omega H) = 0$

Figure 4.2: Empirical moments obtained by 1st-order SIM with macroscopic step H and 1st-order GLA [41] with microscopic step h. Parameters are $\omega = 100$, $c = 0.1$, $\beta = \frac{2c}{\sigma^2} = 10$, $x(0) = 0.8/\omega$, $y(0) = 1.1 + x(0)$, $p_x(0) = 0$, $p_y(0) = 0$; $h = 0.1/\omega$ and H is chosen to be not scaling with ω yet corresponding to a resonant frequency; empirical moments are obtained by averaging 5000 simulations.

1st-order SIM (Integrator 4.2.2) is compared in Figure 4.2 to the benchmark of

Geometric Langevin Integrator (GLA) [41], which is both Boltzmann-Gibbs preserving and path-wise convergent. Agreements on empirical moments of integrated trajectories serve as evidences of convergence in distribution. The large timesteps used by SIM are chosen to be the resonance frequencies and they do produce stable accurate results. $\mathcal{O}(\omega)$-fold acceleration is gained by SIM.

4.2.6 A deterministic numerical example

Consider again the deterministic Fermi-Pasta-Ulam (FPU) problem ([101]; discussed in Section 2.6.3 and illustrated in Figure 2.8), which corresponds to the Hamiltonian:

$$H(q,p) := \frac{1}{2}\sum_{i=1}^{m}(p_{2i-1}^2 + p_{2i}^2) + \frac{\omega^2}{4}\sum_{i=1}^{m}(q_{2i} - q_{2i-1})^2 + \sum_{i=0}^{m}(q_{2i+1} - q_{2i})^4 \quad (4.18)$$

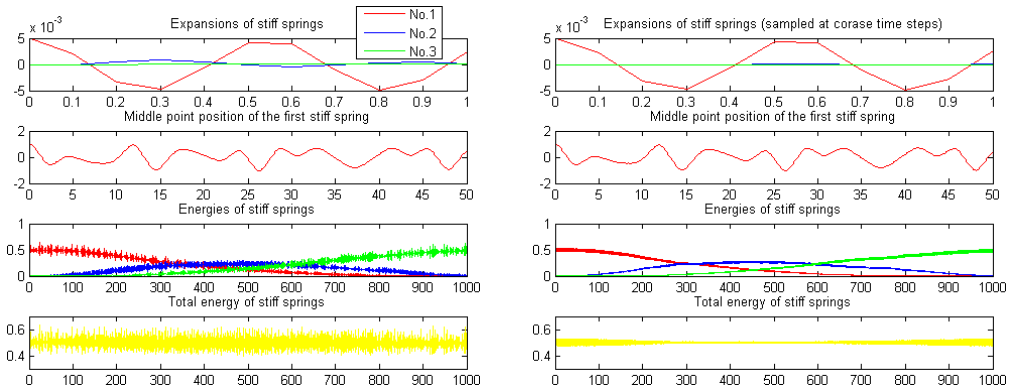
Conventionally, the following transformation is used

$$\begin{cases} x_i &= (q_{2i} + q_{2i-1})/\sqrt{2} \\ x_{m+i} &= (q_{2i} - q_{2i-1})/\sqrt{2} \\ y_i &= (p_{2i} + p_{2i-1})/\sqrt{2} \\ y_{m+i} &= (p_{2i} - p_{2i-1})/\sqrt{2} \end{cases} , i = 1, ...m, \quad (4.19)$$

so that the stiff potential is diagonalized:

$$\begin{cases} H(x,y) &= \frac{1}{2}\sum_{i=1}^{2m} y_i^2 + V_f(x) + V_s(x) \\ V_f(x) &= \frac{\omega^2}{2}\sum_{i=1}^{m} x_{m+i}^2 \\ V_s(x) &= \frac{1}{4}((x_1 - x_{m+1})^4 + \sum_{i=1}^{m-1}(x_{i+1} - x_{m+i+1} - x_i - x_{m+i})^4 + (x_m + x_{2m})^4) \end{cases}$$

We present in Figure 4.3 1st-order SIM simulation (Integrator 4.2.1) together with Variational Euler (also known as symplectic Euler) simulation of FPU over a time span of $\mathcal{O}(\omega)$. Good results are obtained by SIM beyond the timescale of $\mathcal{O}(1)$ (as guaranteed by Theorem 4.2.1) but actually over $\mathcal{O}(\omega)$, and 200-fold ($\omega = 200$) acceleration is gained at the same time.



(a) ***1st-order SIM, large step*** $H = 0.1$

(b) ***Variational Euler, small step*** $h = 0.1/\omega = 0.0005$

Figure 4.3: Simulations of FPU over $T = 5\omega$. Parameters are $\omega = 200$, $m = 3$, $x(0) = [1, 0, 0, 1/\omega, 0, 0]$, $y(0) = [0, 0, 0, 0, 0, 0]$. Different subplots use different time axes to accentuate different timescales: Subplot1 shows scaled expansions of three stiff springs x_{m+i}, which are fast variables; Subplot2 shows scaled middle point position of the first stiff spring x_1, which is one of the slow variables; Subplot3 shows the energy transferring pattern among stiff springs, which is even slower; Subplot4 shows the near-constant total energy of three stiff springs. The fast variables of stiff spring expansions are in fact oscillating much faster than shown in Subplots 1, because Subplots 1 are plotted by interpolating mesh points with a coarse mesh size of H.

Notice that mollified impulse methods with ShortAverage, LongAverage or LinearAverage filters (introduced in [109]) do not accurately capture the rates of energy exchanging among stiff springs over a time span longer than $\mathcal{O}(\omega)$ (results not shown).

4.3 Quasi-quadratic stiff potentials

Most results in this section can be found in a submitted manuscript [273].

4.3.1 The general methodology for arbitrary stiff potentials

Consider the numerical integration of a multiscale Hamiltonian system:

$$\begin{cases} M \begin{bmatrix} \dot{q}^{fast} \\ \dot{q}^{slow} \end{bmatrix} &= \begin{bmatrix} p^{fast} \\ p^{slow} \end{bmatrix} \\ \begin{bmatrix} \dot{p}^{fast} \\ \dot{p}^{slow} \end{bmatrix} &= -\nabla V(q^{fast}, q^{slow}) - \epsilon^{-1} \nabla U(q^{fast}, q^{slow}) \end{cases} \tag{4.20}$$

where q^{slow}, p^{slow} and q^{fast}, p^{fast} are slow and fast degrees of freedom (in the sense that slow degrees of freedom have bounded time derivatives, whereas time derivatives of fast ones may grow unboundedly as $\epsilon \to 0$). Notice that not all stiff Hamiltonian systems are multiscale, and whether a separation of timescales exists depends on specific forms of $V(\cdot)$, $U(\cdot)$ and initial conditions; to the authors' knowledge, however, a generic theory that determines whether a stiff system is multiscale has not been fully developed yet.

Assume without loss of generality that M is the identity matrix. The governing ODE system (4.20) can be written as the sum of three vector fields:

$$\begin{cases} \dot{q}^{fast} &= 0 \\ \dot{p}^{fast} &= 0 \\ \dot{q}^{slow} &= p^{slow} \\ \dot{p}^{slow} &= 0 \end{cases} \quad \begin{cases} \dot{q}^{fast} &= 0 \\ \dot{p}^{fast} &= -\frac{\partial V}{\partial q^{fast}} \\ \dot{q}^{slow} &= 0 \\ \dot{p}^{slow} &= -\frac{\partial V}{\partial q^{slow}} \end{cases} \quad \begin{cases} \dot{q}^{fast} &= p^{fast} \\ \dot{p}^{fast} &= -\epsilon^{-1} \frac{\partial U}{\partial q^{fast}} \\ \dot{q}^{slow} &= 0 \\ \dot{p}^{slow} &= -\epsilon^{-1} \frac{\partial U}{\partial q^{slow}} \end{cases}$$

Denote the exact flow map of each system respectively by $\phi^i(s), i = 1, 2, 3$ over a time of s. It is easy to see that all of them are symplectic.

Observe that ϕ^1 and ϕ^2 are analytically available. We only consider the case where ϕ^3 is also analytically or numerically known; more precisely, the numerical solution $\tilde{\phi}^3$ has to have a consistent uniform local error over a coarse time step $H = o(1)$, i.e., $\|\tilde{\phi}^3(H) - \phi^3(H)\| \leq CH^2$ for a constant C independent of ϵ^{-1}. This can be satisfied for arbitrary $U(\cdot)$ by a symplectic integration with a microscopic timestep $h = o(\sqrt{\epsilon})$, which is in the same spirit as the impulse method. On

the other hand, for specific types of $U(\cdot)$, such as quasi-quadratic stiff potentials (defined in Section 4.3.2), a method alternative to fine-scale integration can be proposed (see Section 4.3.4 and 4.3.5).

Having the three flow maps at hand, one-step update of the proposed method is obtained by composing the three flow maps: $\phi^1(H) \circ \phi^2(H) \circ \phi^3(H)$. Notice that any split can result in a convergent numerical scheme, but this particular split treats two timescales independently and therefore is uniformly convergent at least in the quasi-quadratic stiff potential case (illustrated later); also, it results in a symplectic scheme.

Remark 4.3.1. *If there were no slow variable, we would compose the flows of* $\begin{cases} \dot{q}^{fast} = p^{fast} \\ \dot{p}^{fast} = -\epsilon^{-1} \frac{\partial U}{\partial q^{fast}} \end{cases}$ *and* $\begin{cases} \dot{q}^{fast} = 0 \\ \dot{p}^{fast} = -\frac{\partial V}{\partial q^{fast}} \end{cases}$ *and obtain a first-order version of the original impulse method.*

Remark 4.3.2. *There are also alternative higher-order ways of composing these flow maps; see, for instance, [128, 213]. In fact, the original impulse method is second-order and can be constructed from a second-order composition scheme. However, we will stick to first-order Lie-Trotter ($\phi^1(H) \circ \phi^2(H) \circ \phi^3(H)$) in this section.*

Remark 4.3.3. *If the impulse method were used to integrate* (4.20)*, its practical implementation requires a numerical approximation to the stiff system*

$$\begin{cases} \ddot{q}^{fast} &= -\epsilon^{-1} \partial U / \partial q^{fast}(q^{fast}, q^{slow}) \\ \ddot{q}^{slow} &= -\epsilon^{-1} \partial U / \partial q^{slow}(q^{fast}, q^{slow}) \end{cases}, \tag{4.21}$$

which generally needs to be based on a numerical integration with small steps. The advantage of the impulse method over Verlet is that ∇V only needs to be evaluated at coarse timesteps, but nevertheless its computational cost blows up as $\epsilon \to 0$.

The proposed method, on the other hand, only requires an exact solution or a

numerical approximation to the following stiff system:

$$\begin{cases} \ddot{q}^{fast} &= -\epsilon^{-1}\partial U/\partial q^{fast}(q^{fast},\cdot) \\ \dot{p}^{slow} &= -\epsilon^{-1}\partial U/\partial q^{slow}(q^{fast},\cdot) \end{cases}, \tag{4.22}$$

in which q^{slow} is fixed.

Compared to (4.21) *(required by the impulse method),* (4.22) *is always easier to solve or approximate. In another words, whenever the proposed method requires an introduction of microscopic integration, the impulse method does so as well. On the other hand, when the impulse method needs a microscopic step, the proposed method might still be able to use a macroscopic step, as we will immediately see.*

4.3.2 Quasi-quadratic stiff potentials: Introduction

One case in which the proposed method allows macroscopic steps but the impulse method does not is when $U = \frac{1}{2}[q^{fast}]^T K(q^{slow})q^{fast}$, where K is an arbitrary positive definite d_f-by-d_f symmetric-matrix-valued function. This U represents stiff harmonic oscillators with non-constant but slowly varying frequencies, and we call such potentials quasi-quadratic.

In this case, we show that the exact flow map ϕ^3 can be explicitly computed by exponentiating a matrix. To be practical, of course, we numerically approximate the matrix exponential.

Still, if not handled appropriately, the computational cost of the numerical exponentiation blows up rapidly as ϵ decreases and/or the dimension of the system increases. Furthermore, symplecticity would also be jeopardized by inaccuracies in the numerical exponentiations. In fact, there are various approaches to exponentiate a matrix, including diagonalization, series methods, scaling and squaring, ODE solving, polynomial methods, matrix decomposition methods, and splitting, etc., as comprehensively reviewed in [208]; few of them, however, guarantee the resulting implementation of the proposed method to be symplectic (as it analytically should be), unless the computation is executed to a very high precision at

the cost of losing computational efficiency. For instance, the prevailing method of scaling and squaring, which is used by the MATLAB command 'expm', uses a Padé approximation [141], which is unfortunately not symplectic (see (4.37) and corresponding discussion). Similarly, there is no reason for methods based on matrix decompositions (e.g., diagonalization, QR decomposition, etc.; see for instance [121]) to preserve the geometric structure of symplecticity.

In this section, we propose an integrator well-adapted to high-dimensional systems, which computes the exponentiation in an efficient and symplectic way. Only $\mathcal{O}(n)$ matrix multiplication operations at each coarse time step are needed, where n is a preset small integer at most $\log \epsilon^{-1}$. Although simple in appearance, to guarantee the symplecticity (in all variables) of the resulting method and computational efficiency at the same time is a surprisingly difficult problem, and in fact it is highly nontrivial even when $K(q^{slow})$ is a scalar [176].

In addition to a solution to this problem, we also provide a general method for iteratively exponentiating a slowly varying sequence of (possibly high dimensional) matrices in an efficient way (see Sections 4.3.5 and 4.3.11). This method works for any matrices, and it is not restricted to the numerical integration of (4.20). The preservation of symplecticity associated with these two proposed matrix exponentiation schemes (both suit high-dimensional systems; the first one is in Section 4.3.4) is a core difficulty that we addressed.

4.3.3 When the frequency matrix is diagonal

Given a quasi-quadratic stiff potential, the third split vector field writes as

$$\begin{cases} \dot{q}^{fast} &= p^{fast} \\ \dot{p}^{fast} &= -\epsilon^{-1} K(q^{slow}) q^{fast} \\ \dot{q}^{slow} &= 0 \\ \dot{p}^{slow} &= -\epsilon^{-1} \frac{1}{2} [q^{fast}]^T \nabla K(q^{slow}) q^{fast} \end{cases} \tag{4.23}$$

where the last equation is understood as $\dot{p}_i^{slow} = -\epsilon^{-1}\frac{1}{2}[q^{fast}]^T \partial_i K(q^{slow}) q^{fast}$ for $i = 1, \ldots, d_s$.

The flow of this dynamical system on q^{fast} and p^{fast} is just an exponential map, which in this case corresponds to linear combinations of initial conditions with trigonometric coefficients. For p^{slow}, because q^{slow} (and hence $\nabla K(q^{slow})$) is fixed, one could obtain its exact flow by analytically integrating a quadratic function of trigonometric functions.

When $d_f = 1$, the exact flow map of (4.23) over time H is (letting $\omega = \sqrt{\epsilon^{-1}K(q^{slow})}$):

$$\begin{cases} q^{fast} & \mapsto \cos(\omega H) q^{fast} + \sin(\omega H)/\omega p^{fast} \\ p^{fast} & \mapsto -\omega \sin(\omega H) q^{fast} + \cos(\omega H) p^{fast} \\ q^{slow} & \mapsto q^{slow} \\ p^{slow} & \mapsto p^{slow} - \epsilon^{-1}\frac{1}{2}\nabla K(q^{slow}) \frac{1}{4\omega^3}\big(2\omega(H[p^{fast}]^2 + p^{fast}q^{fast} + \omega^2 H[q^{fast}]^2) \\ & \qquad -2\omega p^{fast} q^{fast}\cos(2\omega H) + (-[p^{fast}]^2 + \omega^2[q^{fast}]^2)\sin(2\omega H)\big) \end{cases} \tag{4.24}$$

where again the last equation is understood as

$$p_i^{slow} \mapsto p_i^{slow} - \epsilon^{-1}\frac{1}{2}\partial_i K(q^{slow}) \frac{1}{4\omega^3}\big(2\omega(H[p^{fast}]^2 + p^{fast}q^{fast} + \omega^2 H[q^{fast}]^2) -2\omega p^{fast} q^{fast}\cos(2\omega H) + (-[p^{fast}]^2 + \omega^2[q^{fast}]^2)\sin(2\omega H)\big) \tag{4.25}$$

When $d_f \geq 2$, the obvious method to obtain the exact flow of (4.23) is based on a diagonalization of K. More precisely, since K is symmetric, we can write $\epsilon^{-1}K(q^{slow}) = \epsilon^{-1}Q(q^{slow})^T D(q^{slow}) Q(q^{slow})$, where $\epsilon^{-1}D(q^{slow}) = \mathrm{diag}[\omega_1^2, \ldots, \omega_{d_f}^2])$. Then

$$\exp\left(\begin{bmatrix} 0 & HI \\ -\epsilon^{-1}HK(q^{slow}) & 0 \end{bmatrix}\right) = \begin{bmatrix} Q^T & 0 \\ 0 & Q^T \end{bmatrix} \exp\left(\begin{bmatrix} 0 & HI \\ -\epsilon^{-1}HD & 0 \end{bmatrix}\right) \begin{bmatrix} Q & 0 \\ 0 & Q \end{bmatrix} = \begin{bmatrix} Q^T & 0 \\ 0 & Q^T \end{bmatrix} \cdot \begin{bmatrix} \mathrm{diag}[\cos(\omega_1 H), \ldots, \cos(\omega_{d_f} H)] & \mathrm{diag}[\sin(\omega_1 H)/\omega_1, \ldots, \sin(\omega_{d_f} H)/\omega_{d_f}] \\ \mathrm{diag}[-\sin(\omega_1 H)\omega_1, \ldots, -\sin(\omega_{d_f} H)\omega_{d_f}] & \mathrm{diag}[\cos(\omega_1 H), \ldots, \cos(\omega_{d_f} H)] \end{bmatrix} \begin{bmatrix} Q & 0 \\ 0 & Q \end{bmatrix} \tag{4.26}$$

A similar (but lengthy) calculation will give the expression of the flow on p^{slow}.

If the diagonalization frame of $K(\cdot)$ is constant, i.e., Q does not depend on q^{slow}, then Q needs to be computed only once throughout the simulation, and then the calculation of the flow on q^{fast} and p^{fast} is dominated by the cost of 2 matrix multiplication operations per coarse step (at expense of $\mathcal{O}({d_f}^{2.376})$ per multiplication by the state-of-art Coppersmith-Winograd algorithm [75]). However, if the frame varies (Q depends on q^{slow}), then diagonalizing K at each time step can offset the gain obtained by the macro-time-stepping of the algorithm. This is especially true if d_f is large. Moreover, errors in numerical diagonalizations may accumulate and deteriorate the symplecticity of ϕ^3.

We address those difficulties by proposing a method, described below, for the numerical integration of (4.23) that is symplectic and that remains computationally tractable in high-dimensional cases (large d_f).

4.3.4 Fast matrix exponentiation for the symplectic integration of the entire system

The proposed approximation of ϕ^3 is based on matrix exponentiation. We will first describe its analytical formulation, and then present an accurate numerical approximation that is both symplectic and computationally cheap.

The first step of our method is based on the following property of matrix exponentials illustrated in [288]: if N and M are constant square matrices of the same dimension, then

$$\exp\left(\begin{bmatrix} -N^T & M \\ 0 & N \end{bmatrix} H\right) = \begin{bmatrix} F_2(H) & G_2(H) \\ 0 & F_3(H) \end{bmatrix} \tag{4.27}$$

with

$$\begin{cases} F_2(H) & = \exp(-N^T H) \\ F_3(H) & = \exp(NH) \\ F_3(H)^T G_2(H) & = \int_0^H \exp(N^T s) M \exp(Ns)\, ds \end{cases} \tag{4.28}$$

This could be proven by solving the ODEs that F_2, F_3, G_2 satisfy (see Lemma 4.3.8 for a more difficult case).

Therefore, ordering coordinates as q^{fast}, p^{fast}, taking $N := \begin{bmatrix} 0 & I \\ -\epsilon^{-1}K(q^{slow}) & 0 \end{bmatrix}$ and $M_i := \begin{bmatrix} \epsilon^{-1}\partial_i K(q^{slow}) & 0 \\ 0 & 0 \end{bmatrix}$ with $i = 1, \ldots, d_s$ which indicates the component of the slow variable, we obtain that if

$$\begin{bmatrix} F_2(H) & G_{2,i}(H) \\ 0 & F_3(H) \end{bmatrix} := \exp\left(\begin{bmatrix} -N^T & M_i \\ 0 & N \end{bmatrix} H \right) \tag{4.29}$$

then the (linear) flow map on q^{fast}, p^{fast} is given by

$$\exp(NH) = F_3(H) \tag{4.30}$$

and the (nonlinear) drift on p^{slow} is given by

$$\begin{aligned} &\int_t^{t+H} q^{fast}(s)^T \epsilon^{-1}\partial_i K(q^{slow}) q^{fast}(s)\, ds \\ &= \int_0^H \begin{bmatrix} q^{fast}(t) \\ p^{fast}(t) \end{bmatrix}^T \exp(N^T s) M_i \exp(Ns) \begin{bmatrix} q^{fast}(t) \\ p^{fast}(t) \end{bmatrix} ds \\ &= \begin{bmatrix} q^{fast}(t) \\ p^{fast}(t) \end{bmatrix}^T F_3(H)^T G_{2,i}(H) \begin{bmatrix} q^{fast}(t) \\ p^{fast}(t) \end{bmatrix} \end{aligned} \tag{4.31}$$

Therefore, $\phi^3(H)$ is given by:

$$\begin{cases} \begin{bmatrix} q^{fast} \\ p^{fast} \end{bmatrix} & \mapsto F_3(H) \begin{bmatrix} q^{fast} \\ p^{fast} \end{bmatrix} \\ q^{slow} & \mapsto q^{slow} \\ p_i^{slow} & \mapsto p_i^{slow} - \frac{1}{2} \begin{bmatrix} q^{fast} \\ p^{fast} \end{bmatrix}^T F_3(H)^T G_{2,i}(H) \begin{bmatrix} q^{fast} \\ p^{fast} \end{bmatrix} \end{cases} \tag{4.32}$$

where in the last equation $i = 1, \ldots, d_s$.

In addition, our specific choice of M_i is a symmetric matrix for each i, because $K(\cdot)$ is symmetric. Consequently, $\exp(N^T s) M_i \exp(Ns)$ is symmetric, and therefore

$$F_3(H)^T G_{2,i}(H) = (F_3(H)^T G_{2,i}(H))^T \tag{4.33}$$

Assuming we have F_3 and $G_{2,i}$ (which will be given by Integrator 4.3.2), (4.20) can be integrated by the following:

Integrator 4.3.1. *Symplectic integrator for* (4.20) *with* $U = \frac{1}{2}[q^{fast}]^T K(q^{slow}) q^{fast}$*: its one-step update mapping* q_k, p_k *onto* q_{k+1}, p_{k+1} *with a macroscopic timestep* H *is given by:*

$$\begin{cases} q^{slow}_{k'} &= q^{slow}_k + H p^{slow}_k \\ q^{fast}_{k'} &= q^{fast}_k \\ p^{slow}_{k'} &= p^{slow}_k - H \partial V / \partial q^{slow}(q^{slow}_{k'}, q^{fast}_{k'}) \\ p^{fast}_{k'} &= p^{fast}_k - H \partial V / \partial q^{fast}(q^{slow}_{k'}, q^{fast}_{k'}) \end{cases} \tag{4.34}$$

$$\begin{cases} \begin{bmatrix} q^{fast}_{k+1} \\ p^{fast}_{k+1} \end{bmatrix} &= F_{3,k} \begin{bmatrix} q^{fast}_{k'} \\ p^{fast}_{k'} \end{bmatrix} \\ q^{slow}_{k+1} &= q^{slow}_{k'} \\ p^{slow}_{k+1,i} &= p^{slow}_{k',i} - \frac{1}{2} \begin{bmatrix} q^{fast}_{k'} \\ p^{fast}_{k'} \end{bmatrix}^T F^T_{3,k} G_{2,k,i} \begin{bmatrix} q^{fast}_{k'} \\ p^{fast}_{k'} \end{bmatrix} \end{cases} \tag{4.35}$$

where $F_{2,k}$, $G_{2,k,i}$ $(i = 1, \ldots, d_s)$ *and* $F_{3,k}$ *are numerical approximations of that in* (4.29) *at each time step* k' *(using* $q^{slow}_{k'}$*), for instance computed by Integrator 4.3.2.*

Remark 4.3.4. *Integrator 4.3.1 could be mollified by using Hamilton-Pontryagin-Marsden principle (Section 4.1) if additional stability is desired. This is one possible future direction.*

To numerically approximate the above flow map (4.32), i.e., to obtain $F_{3,k}$ and $G_{2,k,i}$, we need to ensure two points: (i) an approximation of the matrix

exponential (and hence $F_{3,k}$ and $G_{2,k,i}$) will not affect the symplecticity of the resulting approximation of ϕ^3; (ii) the numerical computation of the exponential (4.29) will not offset the savings gained by using a coarse timestep. This is not easy; in fact, here is an illustration of a popular non-decomposition-based exponentiation method that fails to satisfy this symplecticity condition:

Example: MATLAB function 'expm' [141] uses a scaling and squaring strategy based on the following identity:

$$\exp(X) = [\exp(X/2^n)]^{2^n} \tag{4.36}$$

where n is a big enough preset integer such that $X/2^n$ has a small norm, and therefore Padé approximation [141] could be employed to approximate $\exp(X/2^n)$. The simplest (1,0) Padé approximation, which is essentially Taylor expansion to 1st-order, gives

$$\exp(X) \approx [I + X/2^n]^{2^n} \tag{4.37}$$

However, this approximation is not symplectic. For instance, consider a counterexample of $X = \begin{bmatrix} 0 & I \\ -\Omega^2 & 0 \end{bmatrix}$. Obviously, this corresponds to a vectorial harmonic oscillator, and $\exp(X)$ ought to be symplectic. However, it can be easily checked that $A := I + X/2^n$ does not satisfy $A^T J A = J$ and hence is not symplectic. □

Our idea is to obtain $F_{2,k}$ and $F_{3,k}$ using a modified scaling and squaring strategy, in which the Padé approximation is replaced by a symplectic approximation originated from a reversible symplectic integrator (we use Velocity-Verlet). More precisely, suppose $h > 0$ is a small constant, then we have the following identity:

$$\begin{bmatrix} F_{2,k}(H) & G_{2,k,i}(H) \\ 0 & F_{3,k}(H) \end{bmatrix} = \begin{bmatrix} F_{2,k}(h) & G_{2,k,i}(h) \\ 0 & F_{3,k}(h) \end{bmatrix}^{H/h} \tag{4.38}$$

$F_{3,k}(h)$ can be approximated by the following:

$$\exp\begin{bmatrix} 0 & hI \\ -h\epsilon^{-1}K(q_{k'}^{slow}) & 0 \end{bmatrix} \approx \begin{bmatrix} I - \frac{h^2}{2}\epsilon^{-1}K(q_{k'}^{slow}) & h\left(I - \frac{h^2}{4}\epsilon^{-1}K(q_{k'}^{slow})\right) \\ -h\epsilon^{-1}K(q_{k'}^{slow}) & I - \frac{h^2}{2}\epsilon^{-1}K(q_{k'}^{slow}) \end{bmatrix}, \tag{4.39}$$

which can be easily checked to be symplectic thanks to the specific $\mathcal{O}(h^2)$ and $\mathcal{O}(h^3)$ corrections in the above expression.

It is a classical result (global error bound of Velocity-Verlet) that links $F_{3,k}(H)$ with the approximated $F_{3,k}(h)$:

$$\begin{aligned} &\left\| \exp\begin{bmatrix} 0 & HI \\ -H\epsilon^{-1}K(q_{k'}^{slow}) & 0 \end{bmatrix} - \begin{bmatrix} I - \frac{h^2}{2}\epsilon^{-1}K(q_{k'}^{slow}) & h\left(I - \frac{h^2}{4}\epsilon^{-1}K(q_{k'}^{slow})\right) \\ -h\epsilon^{-1}K(q_{k'}^{slow}) & I - \frac{h^2}{2}\epsilon^{-1}K(q_{k'}^{slow}) \end{bmatrix}^{H/h} \right\|_2 \\ &\qquad \leq \epsilon^{-1}C\exp(CH)h^2 \end{aligned} \tag{4.40}$$

for some constant $C > 0$, because the approximation in (4.39) corresponds to the celebrated Velocity-Verlet integrator with updating rule:

$$\begin{cases} x_{i+\frac{1}{2}} = x_i + \frac{h}{2}y_i \\ y_{i+1} = y_i - h\epsilon^{-1}K(q_{k'}^{slow})x_{i+\frac{1}{2}} \\ x_{i+1} = x_{i+\frac{1}{2}} + \frac{h}{2}y_{i+1} \end{cases} \tag{4.41}$$

for the system $\begin{cases} \dot{x} = y \\ \dot{y} = -\epsilon^{-1}K(q_{k'}^{slow})x \end{cases}$, which is well-known to have a 2nd-order global error.

We can repeat the same procedure to get an approximation of $F_{2,k}(H)$ by using the following approximated $F_{2,k}(h)$:

$$\exp\begin{bmatrix} 0 & h\epsilon^{-1}K^T(q_{k'}^{slow}) \\ -hI & 0 \end{bmatrix} \approx \begin{bmatrix} I - \frac{h^2}{2}\epsilon^{-1}K^T(q_{k'}^{slow}) & h\epsilon^{-1}K^T(q_{k'}^{slow}) \\ -h\left(I - \frac{h^2}{4}\epsilon^{-1}K^T(q_{k'}^{slow})\right) & I - \frac{h^2}{2}\epsilon^{-1}K^T(q_{k'}^{slow}) \end{bmatrix} \tag{4.42}$$

To approximate $G_{2,k,i}(h)$, we follow the result of Lemma 4.3.1 that in the

continuous case $G_{2,k,i} = -J\frac{\partial}{\partial q^{slow}_{k',i}}F_{3,k}$ and let

$$G_{2,k,i}(h) = -J\partial_i F_{3,k}(h) \approx \begin{bmatrix} h\epsilon^{-1}\frac{\partial}{\partial q^{slow}_{k',i}}K(q^{slow}_{k'}) & \frac{h^2}{2}\epsilon^{-1}\frac{\partial}{\partial q^{slow}_{k',i}}K(q^{slow}_{k'}) \\ -\frac{h^2}{2}\epsilon^{-1}\frac{\partial}{\partial q^{slow}_{k',i}}K(q^{slow}_{k'}) & -\frac{h^3}{4}\epsilon^{-1}\frac{\partial}{\partial q^{slow}_{k',i}}K(q^{slow}_{k'}) \end{bmatrix} \tag{4.43}$$

Notice that if (1,0) Padé approximation (i.e., 1st-order Taylor expansion) is used, we will get

$$G_{2,k,i}(h) \approx hM_i = \begin{bmatrix} h\epsilon^{-1}\frac{\partial}{\partial q^{slow}_{k',i}}K(q^{slow}_{k'}) & 0 \\ 0 & 0 \end{bmatrix} \tag{4.44}$$

Naturally, (4.43) is a higher-order correction of this.

$G_{2,k,i}(H)$ will also be accurate: since the accuracy of (4.37) is well established, the higher-order corrections that we add in $F_{2,k}(H), F_{3,k}(H), G_{2,k,i}(H)$ will not lead to a less accurate scheme. This can immediately be seen in the context of the numerical integration of a stable system, where a local error of $\mathcal{O}(h^2)$ will only lead to a global error of at most $\epsilon^{-1}CHh$ [174]. We also refer to Appendix A in [208] for an analogous error analysis if one prefers to directly work with matrices.

To sum up, the following numerical approximation of $F_{3,k}$ and $G_{2,k,i}$ will simultaneously guarantee symplecticity, accuracy, and efficiency:

Integrator 4.3.2. *Matrix exponentiation scheme that complements the updating rule of Integrator 4.3.1. $n \geq 1$ is an integer controlling the accuracy of the approximation of the matrix exponentials. k is the same index as the one used in Integrator 4.3.1, and the following needs to be done for each k:*

1. *Evaluate $K_k := K(q^{slow}_{k'})$ and $\partial_i K_k := \frac{\partial}{\partial q^{slow}_{k',i}}K(q^{slow}_{k'})$. Let $h = H/2^n$,*

$$A_k := \begin{bmatrix} I - \epsilon^{-1}K_k\frac{h^2}{2} & \epsilon^{-1}K_k h \\ -h(I - \epsilon^{-1}K_k\frac{h^2}{4}) & I - \epsilon^{-1}K_k\frac{h^2}{2} \end{bmatrix}, \tag{4.45}$$

$$C_k := \begin{bmatrix} I - \epsilon^{-1} K_k \frac{h^2}{2} & h(I - \epsilon^{-1} K_k \frac{h^2}{4}) \\ -\epsilon^{-1} K_k h & I - \epsilon^{-1} K_k \frac{h^2}{2} \end{bmatrix}, \tag{4.46}$$

and for $i = 1, \ldots, d_s$,

$$B_{k,i} := \begin{bmatrix} \epsilon^{-1} \partial_i K_k h & \epsilon^{-1} \partial_i K_k \frac{h^2}{2} \\ -\epsilon^{-1} \partial_i K_k \frac{h^2}{2} & -\epsilon^{-1} \partial_i K_k \frac{h^3}{4} \end{bmatrix}. \tag{4.47}$$

2. *Let* $F^1_{2,k} := A_k$, $G^1_{2,k,i} := B_{k,i}$, $F^1_{3,k} := C_k$, *then repetitively apply* $\begin{bmatrix} F^{j+1}_{2,k} & G^{j+1}_{2,k,i} \\ 0 & F^{j+1}_{3,k} \end{bmatrix} := \begin{bmatrix} F^{j}_{2,k} & G^{j}_{2,k,i} \\ 0 & F^{j}_{3,k} \end{bmatrix}^2 = \begin{bmatrix} F^j_{2,k} F^j_{2,k} & F^j_{2,k} G^j_{2,k,i} + G^j_{2,k,i} F^j_{3,k} \\ 0 & F^j_{3,k} F^j_{3,k} \end{bmatrix}$ *for* $j = 1, \ldots, n$.
3. *Define* $F_{2,k} := F^{n+1}_{2,k}$, $G_{2,k,i} := G^{n+1}_{2,k,i}$, $F_{3,k} = F^{n+1}_{3,k}$.

Remark 4.3.5. *The trick for an efficient computation is that raising to the* 2^n*th power is computed by* n *self multiplications, which is due to the semi-group property of the exponentiation operation. An obvious upper bound to guarantee accuracy is* $n \leq C \log \epsilon^{-1}$ *(because the error of numerical exponentiation is bounded by* $\epsilon^{-1} C h = \epsilon^{-1} C H / 2^n$*). In all numerical experiments in this section,* $n = 10$ *worked well, which is a value much smaller than* $\log \epsilon^{-1}$*, and this choice of* n *makes the computation cost of the same order as if* K *could be diagonalized by a constant orthogonal matrix.*

Remark 4.3.6. *Observe that, for a finite-time simulation, the cost of computing* ϕ^3 *numerically with microscopic time-steps blows up with a speed of* $\mathcal{O}(\epsilon^{-1})$*, whereas the cost of matrix exponentiations via Integrator 4.3.2 blows up at a maximum speed of* $\mathcal{O}(\log \epsilon^{-1})$.

Theorem 4.3.2 shows that Integrator 4.3.2 not only ensures $F_{2,k}$ and $F_{3,k}$ to be symplectic, but also guarantees a symplectic approximation to ϕ^3 (Eq. 4.32).

Speed-up is obtained because at each step the computation cost is dominated by $2(d_s + 1)n$ matrix production operations (of $d_f \times d_f$ matrices), where n is a

small integer. If the Coppersmith-Winograd algorithm is used to realize the matrix multiplication operation, then the time complexity for exponentiation at each step is $n\mathcal{O}(d_f^{2.376})$ (assuming $d_s = \mathcal{O}(1)$; the problem of matrix exponentiation is less difficult otherwise).

4.3.5 An alternative matrix exponentiation algorithm based on updating

An alternative way to approximate the flow map (4.32) is to use the slowly varying property of K to generate a symplectic update of the exponential computed at the previous step. The main idea of the method is as follows: given a sequence of matrices $\{X_k\}$ that vary slowly, use the approximation

$$\exp(X_k) = [\exp(X_k/2^n)]^{2^n} \approx [\exp(X_{k-1}/2^n)\exp((X_k - X_{k-1})/2^n)]^{2^n} \qquad (4.48)$$

where n is a preset constant. Again, we use the trick of self-multiplication for computing the $2^n th$ power, and efficiency is guaranteed exactly as before.

Accuracy is achieved because, as shown in the following theorem, the approximation error decreases at an exponential rate with respect to n.

Theorem 4.3.1. *Theorem 5 in [208]:*

$$\|\exp(A+B) - (\exp(A/2^n)\exp(B/2^n))^{2^n}\|_2 \leq 2^{-n-1} e^{\max(\mu(A+B),\mu(A)+\mu(B))} \|[A,B]\|_2 \qquad (4.49)$$

where $\mu(X)$ is the maximum eigenvalue of $(X^* + X)/2$, and $[A,B] = AB - BA$ is the canonical Lie bracket.

Remark 4.3.7 (Generality)**.** *This exponentiation method based on corrections* (4.48) *is not limited to the integration of* (4.20)*, but works for repetitive exponentiations of any slowly varying matrix. It would also work for a set of matrices, as long as they could be indexed to ensure a slow variation.*

Remark 4.3.8 (Other possible updating methods)**.** *Regarding updating matrix exponentials, since there are results such as [95] on relationships between perturbed*

eigenvalues and perturbation in the matrix, a natural thought is to use eigenstructures that were explored in the previous step as initial conditions in iterative algorithms (such as Jacobi-Davidson for eigenvalues [259] or Rayleigh Quotient for extreme eigenvalues [283]). This idea, however, did not significantly accelerate the computation as we explored in numerical experiments with an incomplete pool of methods. Other matrix decompositions methods (QR for instance) did not gain much from previous decompositions either in our numerical investigations.

For our purpose of integration, X_k and A are identified with N in Section 4.3.4 at each timestep, and B is identified as the difference in N's between consecutive steps. Since $K(q^{slow})$ (and hence N as well) is changing slowly, $\|B\|_2 \ll \|A\|_2$; furthermore, the calculation of $[A, B]$ (omitted; notice that B is nilpotent) shows that $\|[A, B]\| \ll \|A\|$. Therefore, the error bound here (4.49) is much smaller than that based on scaling and squaring for the same n. Consequently, we will be able to further decrease the value of n by a few (not a lot because a decrease in n exponentially increases the error).

The reason that we do not identify X_k and A with $\begin{bmatrix} -N^T & M_i \\ 0 & N \end{bmatrix}$ is due to a consideration of symplecticity in all variables, because otherwise $G_{2,k,i}$, obtained as the upper-right block of the exponential, will not be exactly the derivative of $F_{3,k}$. Instead, we let $G_{2,k,i} = -J\frac{\partial}{\partial q^{slow}_{k',i}} F_{3,k}$, where $F_{3,k}$ is updated from $F_{3,k-1}$ using (4.48). Taking the derivative, however, incurs additional computation, because $F_{3,k}$ now depends on not only q^{slow}_k but also q^{slow}_{k-1}, and therefore $\partial q^{slow}_{k',i}/\partial q^{slow}_{(k-1)',j}$ has to be computed so that a chain rule applies to facilitate the computation. In the end, the computational saving based on updating the exponentiation becomes less significant due to the extra cost in updating $\partial q^{slow}_{k',i}/\partial q^{slow}_{(k-1)',j}$, but the implementation becomes more convoluted. We leave the details to Section 4.3.11.

4.3.6 Analysis: Symplecticity

For concise writing, we carry out matrix analysis in block forms in this section. Coordinates are ordered as $q^{fast}, p^{fast}, q^{slow}, p^{slow}$, and therefore $\mathbb{J} = \begin{bmatrix} J & 0 \\ 0 & J \end{bmatrix}$ is the coordinate representation of the canonical symplectic 2-form on the full phase space (abusing notations, we use $J := \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$ to represent the symplectic 2-form on both the fast subspace (for q^{fast}, p^{fast}) and the slow subspace (for q^{slow}, p^{slow}); this should not affect the clarity of the analysis). We also recall that a map $x \mapsto \phi(x)$ is symplectic if and only if $\phi'(x)^T \mathbb{J} \phi'(x) = \mathbb{J}$ or $\phi'(x)^T J \phi'(x) = J$ for all x's (depending on whether x represents all variables or only slow or fast variables).

Lemma 4.3.1. *The numerical approximation to ϕ^3 given by* (4.35) *is symplectic on all variables if and only if $F_{3,k}$ is symplectic and, for $i = 1, \ldots, d_s$, $G_{2,k,i} = -J \frac{\partial F_{3,k}}{\partial q^{slow}_{k',i}}$ (note that for a fixed i, $G_{2,k,i}$, $\frac{\partial F_{3,k}}{\partial q^{slow}_{k',i}}$ and J are $d_f \times d_f$ matrices).*

Proof. For conciseness and convenient reading, write $q^{fast}_{k'}$ and $p^{fast}_{k'}$ as q_f and p_f, $\partial / \partial q^{slow}_{k',i}$ as ∂_i, and $G_{2,k,i}$ and $F_{3,k}$ as $G_{2,i}$ and F_3 in this proof.

The Jacobian of the numerical approximation to $\phi^3 : q_{k'}, p_{k'} \mapsto q_{k+1}, p_{k+1}$ given by (4.35) can be computed as:

$$A = \left[\begin{array}{c|ccc|cc} & F_3 & \partial_1 F_3 \begin{pmatrix} q_f \\ p_f \end{pmatrix} & \cdots & \partial_{d_s} F_3 \begin{pmatrix} q_f \\ p_f \end{pmatrix} & \begin{pmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix} \\ \hline \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} & & & & \begin{array}{c|c} I & 0 \end{array} \\ \hline \begin{array}{c} -\begin{pmatrix} q_f^T & p_f^T \end{pmatrix} F_3^T G_{2,1} \\ \vdots \\ -\begin{pmatrix} q_f^T & p_f^T \end{pmatrix} F_3^T G_{2,d_s} \end{array} & & & & \begin{array}{c|c} -* & I \end{array} \end{array}\right] \tag{4.50}$$

where $(*)_{i,j} = \frac{1}{2}[q_f; p_f]^T \partial_j (F_3^T G_{2,i})[q_f; p_f]$, and the 0's in the upper-right block,

the lower-left block, and the lower-right block respectively corresponds to d_f-by-1, 1-by-d_f, and d_s-by-d_s zero matrices. Notice that we have $F_3^T G_{2,i}$ in the lower-left block because $F_3^T G_{2,i}$ is symmetric (their exact values satisfy this because of (4.33), and their numerical approximations satisfy this because of Lemma 4.3.6).

Symplecticity is equivalent to $A^T \mathbb{J} A = \mathbb{J}$, whose left hand side writes out to be

$$A^T\mathbb{J}A = \left(\begin{array}{c|c|c} F_3^T J & G_{2,1}^T F_3 \begin{pmatrix} q_f \\ p_f \end{pmatrix} \cdots G_{2,d_s}^T F_3 \begin{pmatrix} q_f \\ p_f \end{pmatrix} & \begin{pmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix} \\ \hline \begin{matrix} \begin{pmatrix} q_f^T & p_f^T \end{pmatrix} \partial_1 F_3^T J \\ \vdots \\ \begin{pmatrix} q_f^T & p_f^T \end{pmatrix} \partial_{d_s} F_3^T J \end{matrix} & \multicolumn{2}{c}{\multirow{2}{*}{$\begin{array}{c|c} *^T & I \\ \hline -I & 0 \end{array} \times A$}} \\ \cline{1-1} \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} & \end{array}\right)$$

$$= \left(\begin{array}{c|c|c} F_3^T J F_3 + 0 & (F_3^T J \partial_1 F_3 + G_{2,1}^T F_3) \begin{pmatrix} q_f \\ p_f \end{pmatrix} \cdots (F_3^T J \partial_{d_s} F_3 + G_{2,d_s}^T F_3) \begin{pmatrix} q_f \\ p_f \end{pmatrix} & \begin{pmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix} \\ \hline \triangle & \multicolumn{2}{c}{\begin{array}{c|c} \left([q_f; p_f]^T \partial_i F_3^T J \partial_j F_3 [q_f; p_f]\right)_{i=1,\ldots,d_s; j=1,\ldots,d_s} & 0 \\ \hline 0 & 0 \end{array} + \begin{array}{c|c} -*^T+* & I \\ \hline -I & 0 \end{array}} \end{array}\right) \tag{4.51}$$

where $\triangle$ is naturally negative the transpose of the upper-right block because $A^T \mathbb{J} A$ is skew-symmetric for any A.

This is equal to $\mathbb{J}$ if and only if the upper-left block and the bottom-right block are both J and the upper-right block and the bottom-left block are both zero. The requirement on the upper-left block is

$$F_3^T J F_3 = J \tag{4.52}$$

By the arbitrariness of q_f and p_f, the requirement on upper-right and bottom-left blocks translates to:

$$F_3^T J \partial_i F_3 + G_{2,i}^T F_3 = 0 \tag{4.53}$$

which further simplifies to

$$G_{2,i} = -J\partial_i F_3 \tag{4.54}$$

because $F_3^T J\partial_i F_3 = \partial_i(F_3^T J F_3) - \partial_i F_3^T J F_3 = -\partial_i F_3^T J F_3$, F_3 is invertible due to (4.52), and $J^T = -J$.

The bottom-right block needs to be J, and this requirement is equivalent to

$$[q_f; p_f]^T \left(\partial_i F_3^T J \partial_j F_3 + \frac{1}{2}\partial_i (F_3^T G_{2,j}) - \frac{1}{2}\partial_j (F_3^T G_{2,i})\right)[q_f; p_f] = 0 \tag{4.55}$$

By (4.54), the above left hand side rewrites as

$$\begin{aligned}&[q_f; p_f]^T \left(\partial_i F_3^T J \partial_j F_3 - \frac{1}{2}\partial_i F_3^T J \partial_j F_3 - \frac{1}{2} F_3^T J \partial_i \partial_j F_3 + \frac{1}{2}\partial_j F_3^T J \partial_i F_3 + \frac{1}{2} F_3^T J \partial_j \partial_i F_3\right)[q_f; p_f] \\ &= [q_f; p_f]^T \left(\frac{1}{2}\partial_i F_3^T J \partial_j F_3\right)[q_f; p_f] + [q_f; p_f]^T \left(\frac{1}{2}\partial_j F_3^T J \partial_i F_3\right)[q_f; p_f]\end{aligned} \tag{4.56}$$

Since what are summed up above are just two real numbers, the second number remains the same after taking its transpose, which due to $J^T = -J$ yields

$$[q_f; p_f]^T \left(\frac{1}{2}\partial_j F_3^T J \partial_i F_3\right)[q_f; p_f] = -[q_f; p_f]^T \left(\frac{1}{2}\partial_i F_3^T J \partial_j F_3\right)[q_f; p_f] \tag{4.57}$$

Therefore, (4.55) does hold. □

Lemma 4.3.2. *In Integrator 4.3.2, all A_k and C_k are symplectic; moreover, all $F_{2,k}$ and $F_{3,k}$ are symplectic, too.*

Proof. Straightforward computation using (4.45) and (4.46) shows that $A_k^T J A_k = J$ and $C_k^T J C_k = J$. Moreover, since the product of symplectic matrices is symplectic, all $F_{2,k}$ and $F_{3,k}$, being powers of A_k and C_k, are symplectic. □

Lemma 4.3.3. *In Integrator 4.3.2, $A_k^T C_k = I$ (and equivalently $C_k A_k^T = I$) for all k; moreover, $F_{2,k}^T F_{3,k} = I$ (and equivalently $F_{3,k} F_{2,k}^T = I$).*

Proof. Straightforward computation using (4.45) and (4.46) shows that $A_k^T C_k = I$. Therefore, $(A_k A_k)^T C_k C_k = A_k^T I C_k = I$, and by induction $(A_k^{2^n})^T C_k^{2^n} = I$, i.e., $F_{2,k}^T F_{3,k} = I$. □

Lemma 4.3.4. *In Integrator 4.3.2,* $B_{k,i} = -J\frac{\partial}{\partial q^{slow}_{k',i}}C_k$ *for all* k *and* i*, and* $G_{2,k,i} = -J\frac{\partial}{\partial q^{slow}_{k',i}}F_{3,k}$ *for all* k *and* i*.*

Proof. Use the short-hand notation $\partial_i := \frac{\partial}{\partial q^{slow}_{k,i}}$. Straightforward computation using (4.47) and (4.46) shows that $B_{k,i} = -J\partial_i C_k$ for all k and i.

Since $\begin{bmatrix} F_{2,k} & G_{2,k,i} \\ 0 & F_{3,k} \end{bmatrix} = \begin{bmatrix} A_k & B_{k,i} \\ 0 & C_k \end{bmatrix}^{2^n}$ for all i, by induction, it is only necessary to prove that $G_{2,k,i} = -J\partial_i F_{3,k}$ when $n = 1$. In this case, $G_{2,k,i} = A_k B_{k,i} + B_{k,i}C_k$ and $F_{3,k} = C_k C_k$, and the equality can be proved by the following:

Because $B_{k,i} = -J\partial_i C_k$, $C_k^T A_k = I$ (Lemma 4.3.3) and $J = C_k^T J C_k$ (Lemma 4.3.2), we have

$$C_k^T A_k B_{k,i} = -C_k^T J C_k \partial_i C_k \tag{4.58}$$

Since symplectic matrix is nonsingular, this is

$$A_k B_{k,i} = -J C_k \partial_i C_k \tag{4.59}$$

Adding $B_{k,i}C_k = -J\partial_i C_k C_k$, we have

$$A_k B_{k,i} + B_{k,i}C_k = -J\partial_i (C_k C_k) \tag{4.60}$$

Hence, the induction works. □

Lemma 4.3.5. *In Integrator 4.3.2,* $C_k^T B_{k,i} = B_{k,i}^T C_k$ *for all* k *and* i*.*

Proof. This can be shown by straightforward computation using (4.47) and (4.46). □

Lemma 4.3.6. *In Integrator 4.3.2,* $F_{3,k}^T G_{2,k,i} = G_{2,k,i}^T F_{3,k}$ *for all* k *and* i*.*

Proof. By Lemma 4.3.5, $C_k^T B_{k,i} = B_{k,i}^T C_k$ for all k and i. By Lemma 4.3.3, $A_k^T C_k = I$ and $C_k^T A_k = I$.

Since $\begin{bmatrix} F_{2,k} & G_{2,k,i} \\ 0 & F_{3,k} \end{bmatrix} = \begin{bmatrix} A_k & B_{k,i} \\ 0 & C_k \end{bmatrix}^{2^n}$ for all i, by induction, it is only necessary to prove that $F_{3,k}^T G_{2,k,i} = G_{2,k,i}^T F_{3,k}$ when $n = 1$. In this case, $G_{2,k,i} = A_k B_{k,i} +$

$B_{k,i}C_k$ and $F_{3,k} = C_kC_k$, and this equality can be proved upon observing for all i:

$$C_k^T C_k^T (A_k B_{k,i} + B_{k,i} C_k) = C_k^T B_{k,i} + C_k^T C_k^T B_{k,i} C_k = B_{k,i}^T C_k + C_k^T B_{k,i}^T C_k C_k$$
$$= B_{k,i}^T A_k^T C_k C_k + C_k^T B_{k,i}^T C_k C_k = (A_k B_{k,i} + B_{k,i} C_k)^T C_k C_k \tag{4.61}$$

□

Theorem 4.3.2. *The proposed method (Integrator 4.3.1+4.3.2) is symplectic on all variables.*

Proof. By Lemma 4.3.2, 4.3.4, 4.3.1, and 4.3.6, the numerical approximation to ϕ^3 given by (4.35) is symplectic on all variables.

The flow given by (4.34) is symplectic on all variables as well, because it is the composition of ϕ^1 and ϕ^2, which respectively correspond to Hamiltonians $\mathcal{H}_1(q^{fast}, p^{fast}, q^{slow}, p^{slow}) = \frac{[p^{slow}]^2}{2}$ and $\mathcal{H}_2(q^{fast}, p^{fast}, q^{slow}, p^{slow}) = V(q^{fast}, q^{slow})$, and hence both are symplectic.

Consequently, the proposed method, which composes (4.34) and (4.35), is symplectic. □

4.3.7 Analysis: Uniform convergence

This integrator is convergent due to splitting theory [285], i.e., the global error on $q^{slow}, q^{fast}, p^{slow}, p^{fast}$ is bounded by $\epsilon^{-1}CH$ for some constant $C > 0$ in Euclidean norm.

Moreover, this integrator is uniformly convergent in q under typical or reasonable assumptions, and hence H can be chosen independently from ϵ for stable and accurate integration.

Condition 4.3.1. *We will prove a uniform bound of the global error on position for Integrator 4.3.1 under the following (classical) conditions:*

1. *Regularity: In the integration domain of interest, $\nabla V(\cdot)$ is bounded and Lipschitz continuous with coefficient L, i.e. $\|\nabla V(a) - \nabla V(b)\|_2 \leq L\|a - b\|_2$.*

2. *Stability and bounded energy: For a fixed T and $t < T$, denote by $x(t) = (q(t), p(t))$ the exact solution to (4.20), and by $x_t = (q_t, p_t)$ the discrete numerical trajectory given by Integrator 4.3.1, then $\|x(t)\|_2^2 \leq C$, $\|x_t\|_2^2 \leq C$, $|\mathcal{H}(q(t), p(t))| \leq C$ and $|\mathcal{H}(q_t, p_t)| \leq C$ for some constant C independent of ϵ^{-1} but dependent on initial condition $\| \begin{bmatrix} q_0 \\ p_0 \end{bmatrix} \|_2^2$ and possibly T as well.*

Condition 4.3.2 (Slowly varying frequencies)**.** *Consider the solution $q(s), p(s)$ up to time $s <= H$ to the system*

$$\begin{cases} dq^{fast} &= p^{fast}dt \\ dq^{slow} &= p^{slow}dt \\ dp^{fast} &= -\partial V/\partial q^{fast}(q^{fast}, q^{slow})dt - \epsilon^{-1}K(q^{slow})q^{fast}dt \\ dp^{slow} &= -\partial V/\partial q^{slow}(q^{fast}, q^{slow})dt - \epsilon^{-1}\frac{1}{2}[q^{fast}]^T\nabla K(q^{slow})q^{fast}dt \end{cases}, \tag{4.62}$$

with initial condition $q(0), p(0)$ in the domain of interest that satisfies bounded energy. Assume that q^{fast} can be written as

$$\boldsymbol{Q}(t)\sum_{i=1}^{d_f}\vec{\boldsymbol{e}}_i\sqrt{\epsilon}a_i(t)\cos[\sqrt{\epsilon^{-1}}\theta_i(t)+\phi_i] \tag{4.63}$$

where $\boldsymbol{Q}(t)$ is a slowly varying matrix (i.e., $Q_{ij}(t) \in C^1([0, H])$ and there exists a C independent of ϵ^{-1} such that $\|\boldsymbol{Q}(t)\| \leq C$ and $\|\dot{\boldsymbol{Q}}(t)\| \leq C$ for all $t \in [0, H]$), indicating a slowly varying diagonalization frame, d_f is the dimension of the fast variable, $\vec{\boldsymbol{e}}_i$ are standard vectorial basis of $\mathbb{R}^{d_f}$, $a_i(t)$'s are slowly varying amplitudes (in the same sense as for $\boldsymbol{Q}(t)$), $\theta_i(t)$'s are non-decreasing and slowly varying in the sense that $\theta_i(t) \in C^2([0, H])$, $|\ddot{\theta}_i(t)| \leq C$, $|\theta_i(t)| \leq C$, and $C_1 \leq \dot{\theta}_i(t) \leq C_2$ for some $C > 0, C_1 > 0, C_2 > 0$ independent of ϵ^{-1}, and ϕ_i's are such that $\theta_i(0) = 0$.

Remark 4.3.9. *In the case of constant frequencies ($K(\cdot)$ being a constant) and no slow drift ($V(\cdot)$ being a constant), we have $q^{fast} = \boldsymbol{Q}\sum_{i=1}^{d_f}\vec{\boldsymbol{e}}_i\sqrt{\epsilon}a_i\cos[\sqrt{\epsilon^{-1}}\omega_i t+\phi_i]$ (the amplitude is $\mathcal{O}(\sqrt{\epsilon})$ because of bounded energy). When K is not a constant, Condition 4.3.2 is supported by an asymptotic expansion of q^{fast}. In particular,*

to the leading order in ϵ, we have $\dot{\theta}_i(t) = \omega_i(t)$ where the $\omega_i^2(t)$ are the eigenvalues of $K(q_s^{slow})$. The rigorous justification of this asymptotic expansion for $d_f > 1$ is beyond the scope of this section.

Lemma 4.3.7. *If Condition 4.3.2 holds, there exists $C_1 > 0, C_2 > 0$ independent of ϵ^{-1} such that*

$$\| \int_0^H f(t) q^{fast}(t) dt \| \leq \epsilon \left(C_1 \max_{0 \leq s \leq H} \|f(s)\| + C_2 H \max_{0 \leq s \leq H} \|\dot{f}(s)\| + \mathcal{O}(H^2) \right) \tag{4.64}$$

for arbitrary matrix valued function $f \in C^1([0, H])$ that satisfies $f(0) = 0$.

Proof. Recall the form of q^{fast} in Condition 4.3.2. It is sufficient to prove that for all i's the i-th component of q^{fast} satisfies (4.64), whereas the i-th component writes as:

$$\sqrt{\epsilon} \sum_{j=1}^{d_f} Q_{ij}(t) a_j(t) \cos[\sqrt{\epsilon^{-1}} \theta_i(t) + \phi_i] \tag{4.65}$$

Furthermore, since summation commutes with integral and therefore will only introduce a factor of d_f on the bound, it is sufficient to prove (4.64) for $q^{fast} = \sqrt{\epsilon} Q_{ij}(t) a_j(t) \cos[\sqrt{\epsilon^{-1}} \theta_i(t) + \phi_i]$. On this token, we could assume that we are in the 1D case and absorb $Q(t)$ into $a_j(t)$.

Similarly, slowly varying $a_i(t)$ can be absorbed into the test function $f(t)$, and doing so will only change the constants on the right hand side. Therefore, it will be sufficient to prove that:

$$\left| \int_0^H \sqrt{\epsilon} \cos[\sqrt{\epsilon^{-1}} \theta(t) + \phi] f(t) dt \right| \leq \epsilon \left(C_1 \max_{0 \leq s \leq H} |f(s)| + C_2 H \max_{0 \leq s \leq H} |f'(s)| + \mathcal{O}(H^2) \right) \tag{4.66}$$

for a scalar valued function $f \in C^1([0, H])$ that satisfies $f(0) = 0$.

By Condition 4.3.2, θ is strictly increasing. If we write $\tau = \theta(t)$, there will be a θ^{-1} such that $t = \theta^{-1}(\tau)$. With time transformed to the new variable τ, the integral on the left hand side of (4.66) is equal to

$$\int_0^{\theta(H)} \sqrt{\epsilon} \cos[\sqrt{\epsilon^{-1}} \tau + \phi] f(\theta^{-1}(\tau)) \frac{d\theta^{-1}}{d\tau}(\tau) \, d\tau \tag{4.67}$$

By integration by parts, this is (since $f(0) = 0$)

$$-\epsilon \sin[\sqrt{\epsilon^{-1}}H+\phi]f(H)\frac{1}{\dot{\theta}(H)}+\epsilon\int_0^{\theta(H)} \sin[\sqrt{\epsilon^{-1}}\tau+\phi]\left[\frac{df}{dt}\left(\frac{d\theta^{-1}}{d\tau}\right)^2 + f(\theta^{-1}(\tau))\frac{d^2\theta^{-1}}{d\tau^2}(\tau)\right] \tag{4.68}$$

Because $\ddot{\theta} \leq C$, $\omega - CH \leq \dot{\theta} \leq \omega + CH$, where $\omega := \dot{\theta}(0) \geq C_1 > 0$. Together with $\frac{d\theta^{-1}}{d\tau} = \frac{1}{\dot{\theta}}$, we have $\frac{d\theta^{-1}}{d\tau} = 1/\omega + \mathcal{O}(H)$. Similarly, we also have

$$\frac{d^2\theta^{-1}}{d\tau^2} = \frac{d}{d\tau}\frac{1}{\dot{\theta}(t)} = \frac{dt}{d\tau}\frac{d}{dt}\frac{1}{\dot{\theta}(t)} = -\frac{1}{\dot{\theta}(t)^3}\ddot{\theta}(t) = \mathcal{O}(1) \tag{4.69}$$

It is easy to show that $\theta(H) = \mathcal{O}(H)$. Together with $\sin(\cdot)$ being $\mathcal{O}(1)$, the left hand side in (4.66) is bounded by

$$\begin{aligned}&\epsilon f(H)\mathcal{O}(1) + \epsilon\mathcal{O}(H)\left(\mathcal{O}(1)\max_{0\leq s\leq H}|\dot{f}(s)| + \mathcal{O}(1)\max_{0\leq s\leq H}|f(s)|\right)\\ &\leq \epsilon\left(\mathcal{O}(1)\max_{0\leq s\leq H}|f(s)| + \mathcal{O}(H)\max_{0\leq s\leq H}|\dot{f}(s)|\right)\end{aligned} \tag{4.70}$$

□

Theorem 4.3.3. *If Conditions 4.3.1 and 4.3.2 hold, the proposed method (Integrator 4.3.1) for system* (4.20) *has a uniform global error of* $\mathcal{O}(H)$ *in* q*, given a fixed total simulation time* $T = NH$*:*

$$\|q(T) - q_T\|_2 \leq CH \tag{4.71}$$

where $q(T), p(T)$ *is the exact solution and* q_T, p_T *is the numerical solution;* C *is a positive constant independent of* ϵ^{-1} *but dependent on simulation time* T*, scaleless elasticity matrix* K*, slow potential energy* $V(\cdot)$ *and initial condition* $\| \begin{bmatrix} q_0 \\ p_0 \end{bmatrix} \|_2$.

Proof. Let $\tilde{K}$ be a constant matrix and consider the following system:

$$\begin{cases} d\tilde{q}^{fast} &= \tilde{p}^{fast}dt \\ d\tilde{q}^{slow} &= \tilde{p}^{slow}dt \\ d\tilde{p}^{fast} &= -\partial V/\partial q^{fast}(\tilde{q}^{fast},\tilde{q}^{slow})dt - \epsilon^{-1}\tilde{K}\tilde{q}^{fast}dt \\ d\tilde{p}^{slow} &= -\partial V/\partial \tilde{q}^{slow}(\tilde{q}^{fast},\tilde{q}^{slow})dt \end{cases}, \tag{4.72}$$

Integrator 4.3.1, applied to the system (4.72) under Condition 4.3.1, has been shown in Theorem 4.2.2 to be uniformly convergent in 'scaled energy norm' (Definition 4.2.1, or equivalently, uniformly convergent on position and non-uniformly convergent on momentum). Recall that the 'scaled energy norm' was defined to be

$$\|[\tilde{q},\tilde{p}]\|_E = \sqrt{\tilde{q}^T\tilde{q} + \epsilon\tilde{p}^T\tilde{K}^{-1}\tilde{p}}, \tag{4.73}$$

but in fact $\tilde{K}^{-1}$ is not important because it is just $\mathcal{O}(1)$, and the following definition would also work for the proof there:

$$\|[\tilde{q},\tilde{p}]\|_E = \sqrt{\tilde{q}^T\tilde{q} + \epsilon\tilde{p}^T\tilde{p}} \tag{4.74}$$

Observe that, (4.73) is proportional to the square root of the physical energy. That is why the name.

The system considered here, however, is (4.62). To prove uniform convergence for (4.62), it is sufficient to show that (i) a δ difference between two trajectories of (4.72) in scaled energy norm leads to a difference of $\delta(1+CH)$ in scaled energy norm after a time step H (ii) trajectories of (4.72) and (4.62) starting at the same point remain at at a distance at most $\mathcal{O}(H^2)$ in scaled energy norm after time H, i.e., a 2nd-order uniform local error. (i) was shown by Lemma A.3.5 in Appendix A.3, and we will now prove (ii).

We can assume without loss of generality that we start at time 0, and let $\tilde{K} = K(q^{slow}(0))$, $q^{fast,slow}(0) = \tilde{q}^{fast,slow}(0)$ (where $q^{fast,slow} = (q^{fast}, q^{slow})$) and $p^{fast,slow}(0) = \tilde{p}^{fast,slow}(0)$. We first let $x = \tilde{q}^{fast} - q^{fast}$ and $y = \tilde{p}^{fast} - p^{fast}$, and proceed to bound x and y:

The evolutions of x and y follow from

$$\begin{cases} \dot{x} &= y \\ \dot{y} &= -\left(\frac{\partial V}{\partial q^{fast}}(\tilde{q}) - \frac{\partial V}{\partial q^{fast}}(q)\right) - \epsilon^{-1}\left(\tilde{K}\tilde{q}^f - K(q^{slow})q^{fast}\right) \end{cases} \tag{4.75}$$

Writing $f_1 = -\left(\frac{\partial V}{\partial q^{fast}}(\tilde{q}) - \frac{\partial V}{\partial q^{fast}}(q)\right)$ and $f_2 = (\tilde{K} - K(q^{slow}))q^{fast}$, we have

$$\begin{cases} \dot{x} &= y \\ \dot{y} &= f_1 - \epsilon^{-1}\tilde{K}x - \epsilon^{-1}f_2 \end{cases} \tag{4.76}$$

If we let $B(t) = \exp\left(\begin{bmatrix} 0 & I \\ -\epsilon^{-1}\tilde{K} & 0 \end{bmatrix} t\right)$, we will have

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = B(t)\begin{bmatrix} x(0) \\ y(0) \end{bmatrix} + \int_0^t B(t-s)\begin{bmatrix} 0 \\ f_1 - \epsilon^{-1}f_2 \end{bmatrix} ds \tag{4.77}$$

The first term on the right hand side drops off because $x(0) = 0$ and $y(0) = 0$ by definition.

Since $\tilde{K}$ is a constant matrix, it is sufficient to diagonalize it and treat each diagonal element individually. Hence, assume without loss of generality that we are in the 1D case. Then $B(s) = \begin{bmatrix} \cos(\sqrt{\epsilon^{-1}\tilde{K}}s) & \sin(\sqrt{\epsilon^{-1}\tilde{K}}s)/\sqrt{\epsilon^{-1}\tilde{K}} \\ -\sqrt{\epsilon^{-1}\tilde{K}}\sin(\sqrt{\epsilon^{-1}\tilde{K}}s) & \cos(\sqrt{\epsilon^{-1}\tilde{K}}s) \end{bmatrix}$. As a consequence,

$$y(t) = \int_0^t \cos[\sqrt{\epsilon^{-1}\tilde{K}}(t-s)]\left[f_1 - \epsilon^{-1}(\tilde{K} - K(q^{slow}))q^{fast}\right] ds \tag{4.78}$$

By Lipschitz continuity of ∇V (Item 1 of Condition 4.3.1), we will have

$$|f_1(t)| \leq L|x(t)| = L\left|\int_0^t y(s)ds\right| = \mathcal{O}(t) \tag{4.79}$$

The first inequality holds because f_1 is the difference between partial derivatives of V, which could be bounded by the difference between full derivatives. The last

equality holds because $y = p - \tilde{p}$ is bounded due to the fact that $[q(s), p(s)]$ and $[\tilde{q}(s), \tilde{p}(s)]$ are bounded (Item 2 of Condition 4.3.1). Consequently, we have

$$\left| \int_0^t \cos[\sqrt{\epsilon^{-1}\tilde{K}}(t-s)] f_1 \, ds \right| \leq \int_0^t |f_1| = \mathcal{O}(t^2) \tag{4.80}$$

In order to bound $\int_0^t \cos[\sqrt{\epsilon^{-1}\tilde{K}}(t-s)] \left[\epsilon^{-1}(\tilde{K} - K(q^{slow}))q^{fast}\right] ds$, we use Lemma 4.3.7 (with the choice of $f = \tilde{K} - K(q^{slow})$). Indeed, $\cos[\sqrt{\epsilon^{-1}\tilde{K}}(t-s)]$ can be absorbed into $q^{fast}(s) = \sqrt{\epsilon}\cos[\sqrt{\epsilon^{-1}}\theta(s) + \phi]$: due to an equality $2\cos(A)\cos(B) = \cos(A+B) + \cos(A-B)$, θ will be just added by $\pm\sqrt{\tilde{K}}$ and ϕ will have a new constant value, neither of which will violate Condition 4.3.2.

For f, we clearly have $f = 0$ at $s = 0$. By mean value theorem, there is a ξ_s such that $f(s) = K \circ q^{slow}(0) - K \circ q^{slow}(s) = \frac{dK \circ q^{slow}}{dt}(\xi_s) \cdot s$, and therefore $f(s) = \mathcal{O}(s)$. Similarly, $\dot{f}(s) = \mathcal{O}(1)$. Plotting these two bounds in Lemma 4.3.7, we obtain

$$\left| \int_0^t \cos[\sqrt{\epsilon^{-1}\tilde{K}}(t-s)] \left[\epsilon^{-1}(\tilde{K} - K(q^{slow}))q^{fast}\right] ds \right| = \mathcal{O}(t) \tag{4.81}$$

Putting this together with (4.80), we arrive in $y(t) = \mathcal{O}(t)$, and $x(t) = \int_0^t y(s)\, ds = \mathcal{O}(t^2)$ follows.

Next, we bound y: since

$$\begin{aligned} &\left| \int_0^t \cos[\sqrt{\epsilon^{-1}\tilde{K}}(t-s)] \left[\epsilon^{-1}(\tilde{K} - K(q^{slow}))q^{fast}\right] ds \right| \\ &\qquad = \left| \int_0^t \cos[\ldots]\epsilon^{-1}\mathcal{O}(s)\sqrt{\epsilon}\mathcal{O}(1)\cos[\ldots]\, ds \right| = \epsilon^{-1/2}\mathcal{O}(t^2) \end{aligned} \tag{4.82}$$

we have $y(t) = \epsilon^{-1/2}\mathcal{O}(t^2)$. Together with $x(t) = \mathcal{O}(t^2)$, this is equivalent to $\|[x, y]\|_E = \mathcal{O}(t^2)$.

Similarly, we can bound $q^{slow} - \tilde{q}^{slow}$ and $p^{slow} - \tilde{p}^{slow}$. Let $x^s = q^{slow} - \tilde{q}^{slow}$

and $y^s = p^{slow} - \tilde{p}^{slow}$, then we have:

$$\begin{cases} \dot{x}^s &= y^s \\ \dot{y}^s &= -\left(\frac{\partial V}{\partial q^{slow}}(\tilde{q}) - \frac{\partial V}{\partial q^{slow}}(q)\right) - \epsilon^{-1}\frac{1}{2}[q^{fast}]^T \nabla K(q^{slow}) q^{fast} \end{cases} \tag{4.83}$$

Analogous to before, the first term on the right hand side of the y^s dynamics is $\mathcal{O}(t)$. Since $q^{fast} = \mathcal{O}(\epsilon^{1/2})$, the second term on the right hand side is $\mathcal{O}(1)$. Therefore, $\dot{y}^s = \mathcal{O}(1)$, $y^s(t) = y^s(0) + \mathcal{O}(t) = \mathcal{O}(t)$, and $x^s(t) = x^s(0) + \int_0^t y^s(s)\, ds = \mathcal{O}(t^2)$. For our purpose of fast integration, we use a big timestep $H \geq \sqrt{\epsilon}$, and hence $y^s(H) = \mathcal{O}(H) \leq \epsilon^{-1/2}\mathcal{O}(H^2)$ (notice that if $H < \sqrt{\epsilon}$, we do not even need to prove uniform convergence, because the non-uniform error bound that is guaranteed by Lie-Trotter splitting theory is already very small).

$\mathcal{O}(H^2)$ and $\epsilon^{-1/2}\mathcal{O}(H^2)$ bounds on separations of slow position and slow momentum imply a $\mathcal{O}(t^2)$ uniform bound in scaled energy norm (analogous to that of the fast degrees of freedom). This demonstrates a 2nd-order uniform local error on all variables in scaled energy norm, and therefore concludes the proof. □

Remark 4.3.10. *Unlike* (4.71)*, a global bound on the error of momentum will not be uniform. The error propagation is quantified in scaled energy norm, and in 2-norm we will only have $\epsilon^{-1/2}\mathcal{O}(H^2)$ local error and $\epsilon^{-1/2}\mathcal{O}(H)$ global error on momentum. In fact, Integrator 4.3.1 applied to the constant frequency system* (4.72) *is non-uniformly convergent on momentum (Theorem 4.2.2).*

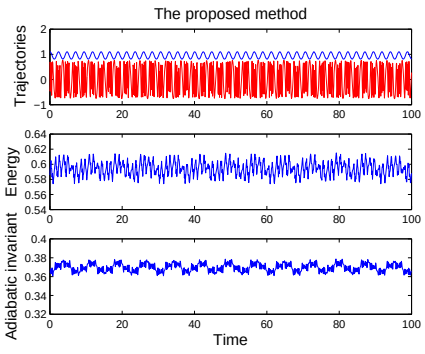
4.3.8 Numerical example: A diagonal frequency matrix

Consider the Hamiltonian example introduced in [176]:

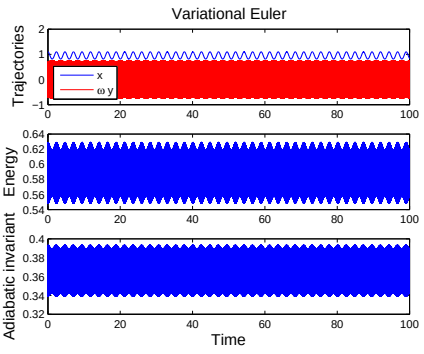
$$\mathcal{H} = \frac{1}{2}p_x^2 + \frac{1}{2}p_y^2 + (x^2 + y^2 - 1)^2 + \frac{1}{2}(1 + x^2)\omega^2 y^2 \tag{4.84}$$

When $\omega = \epsilon^{-1/2} \gg 1$, bounded energy translates to initial conditions $x(0) \sim \omega y(0)$, which satisfy separation of timescales: x is the slow variable, and y is the fast. $K(x) = 1 + x^2$ is trivially diagonal. In addition to conservation of total

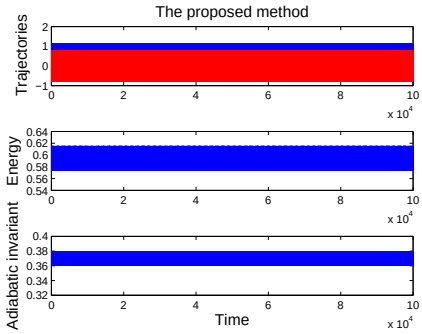
energy, $I = \frac{p_y^2}{2\sqrt{1+x^2}} + \frac{\sqrt{1+x^2}\omega^2 y^2}{2}$ is an adiabatic invariant.



(a) ***The proposed method with coarse timestep*** $H = 0.1$



(b) ***Variational Euler with small timestep*** $h = 0.1/\omega = 0.001$



(c) ***Very long time simulation by the proposed method with coarse timestep*** $H = 0.1$

Figure 4.4: Simulations of a diagonal fast frequency example (4.84) by the proposed method and Variational Euler. $\omega = 100$; $x(0) = 1.1$, $y(0) = 0.7/\omega$.

A comparison between Variational Euler and the proposed method is shown in Figure 4.4. There it can be seen that preservations of energy and adiabatic invariant are numerically captured at least to a very large timescale. Since there is no overhead spent on matrix exponentiation here, an accurate 100x speed up is achieved by the proposed method (because $H/h = 100$).

It is known that the impulse method and its derivatives (such as mollified impulse methods) are not stable if the integration step falls in resonance inter-

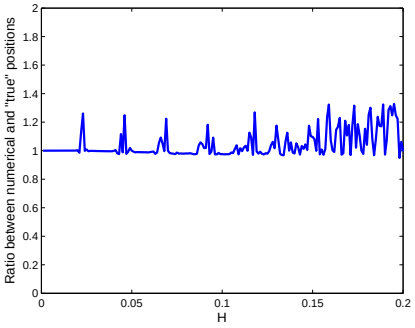


Figure 4.5: Investigation on resonance frequencies of the proposed method on example (4.84). The ratio between $x(T)|_{T=100}$ integrated by the proposed method integration and benchmark provides the ruler: a ratio closer to 1 means a more accurate integration, and deviations from 1 stand for step lengths that correspond to resonance frequencies. Time step H samples from 0.001 to 0.2 with an increment of 0.001. $\omega = 100$; $x(0) = 1.1$, $y(0) = 0.7/\omega$. Benchmark is obtained by fine VE integration with $h = 0.01/\omega$.

vals (mollified impulse methods have much narrower resonance intervals, which however still exist) [109, 55]. Similarly, it will be very unnatural if the proposed method does not have resonance, because it reduces to a 1st-order version of impulse methods when there is no slow variable (Remark 4.3.1). In fact, in our numerical investigation (Figure 4.5), we clearly observe resonance frequencies before the integration step reaches the unstable limit (around $H \approx 0.5$), and widths of resonant intervals increase as H grows for this particular example; however, we will not carry out a systematic analysis on resonance.

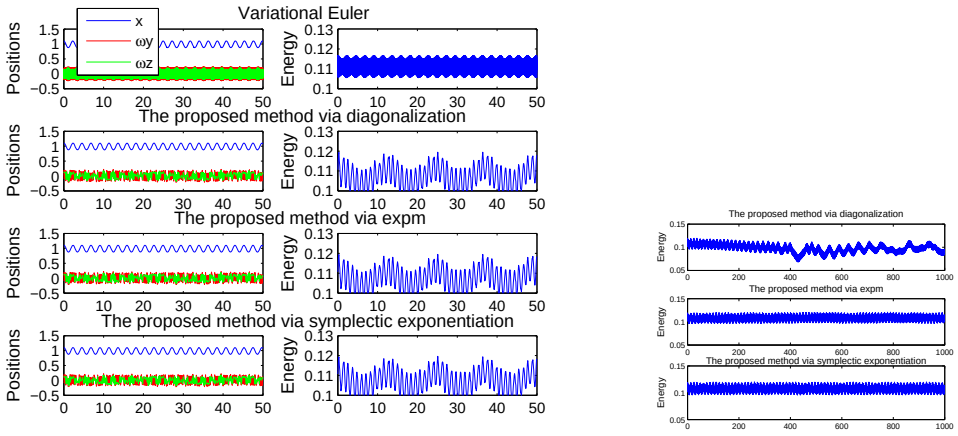
4.3.9 Numerical example: A non-diagonal frequency matrix

Extend the previous example to a toy example of 3 degrees of freedom:

$$\mathcal{H} = \frac{1}{2}p_x^2+\frac{1}{2}p_y^2+\frac{1}{2}p_z^2+(x^2+y^2+z^2-1)^2+\frac{1}{2}\omega^2 \begin{bmatrix} y \\ z \end{bmatrix}^T \begin{bmatrix} 1+x^2 & x^2-1 \\ x^2-1 & 3x^2 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} \quad (4.85)$$

It is easy to check that eigenvalues of $K(x) = \begin{bmatrix} 1+x^2 & x^2-1 \\ x^2-1 & 3x^2 \end{bmatrix}$ are both positive when $x > 0.44$, which will always be true if the initial condition of x

stays close to 1 and ω is big enough. In this case, bounded energy again implies $x(0) \sim \omega y(0) \sim \omega z(0)$ and gives clear separation of timescales: x is the slow variable and y and z are the fast. $K(x)$ has its orthogonal frame for diagonalization as well as its eigenvalues slowly varying with time.



(a) ***Till time 50*** **(b)** ***Till time 1000***

Figure 4.6: Simulations of a non-diagonal fast frequency example (4.85) by Variational Euler, the proposed method with different implementations of matrix exponentiations. $\omega = 100$, VE uses $h = 0.1/\omega = 0.001$ and the proposed method uses $H = 0.1$ and $n = 10$; $x(0) = 1.1$, $y(0) = 0.2/\omega$, $z(0) = 0.1/\omega$, and initial momenta are zero.

Figure 4.6 shows a comparison between Variational Euler, the proposed method with the matrix exponentiations computed by diagonalization and analytical integration (Eq. 4.26; diagonalization implemented by MATLAB command 'diag'), and the proposed methods based on exponentiations (Eq. 4.27 and 4.32) via MATLAB command 'expm' [141] and via the fast matrix exponentiation method (Integrator 4.3.2). The default MATLAB matrix multiplication operation is used. All implementations of the proposed method are accurate, except that numerical errors in repetitive diagonalizations contaminated the symplecticity of the corresponding implementation over a long time simulation (as suggested by drifted energy), whereas two other implementations, respectively based on accurate but slow 'expm' and fast symplectic exponentiations, do not have this issue. In a typical notebook run with MATLAB R2008b, the above four methods respectively spent

11.12, 0.23, 0.29 and 0.24 seconds on the same integration (till time 50), while 0, 0.14, 0.18, and 0.14 seconds were spent on matrix exponentiations. Computational gain by the symplectic exponentiation algorithm will be much more significant as the fast dimension becomes higher. Notice also that the computational gain by the proposed method over Variational Euler will go to infinity as $\epsilon \to 0$, even if the fast matrix exponentiation method is not employed.

4.3.10 Numerical example: A high-dimensional non-diagonal frequency matrix

Consider an arbitrarily high-dimensional example:

$$\mathcal{H} = \frac{1}{2}p^2 + \frac{1}{2}y^T y + (x^T x + q^2 - 1)^2 + \frac{1}{2}\omega^2 x^T T(q) x \tag{4.86}$$

where $q, p \in \mathbb{R}$ correspond to the slow variable, $x, y \in \mathbb{R}^{d_f}$ correspond to fast variables, and $T(q)$ is the following Toeplitz matrix valued function:

$$T(q) = \begin{bmatrix} 1 & \hat{q}^1 & \hat{q}^2 & \dots & \hat{q}^{d_f-1} \\ \hat{q}^1 & 1 & \hat{q}^1 & \dots & \hat{q}^{d_f-2} \\ \hat{q}^2 & \hat{q}^1 & 1 & \dots & \hat{q}^{d_f-3} \\ & & \vdots & & \\ \hat{q}^{d_f-1} & \hat{q}^{d_f-2} & \hat{q}^{d_f-3} & \dots & 1 \end{bmatrix} \tag{4.87}$$

where $\hat{q} = q/2$ so that eigenvectors and eigenvalues vary slowly with q given an initial condition of $q(0) \approx 1$. Note that the expression of $T(\cdot)$ is highly nonlinear.

We present in Figure 4.7 a comparison between Variational Euler and the proposed methods with the matrix exponentials computed by MATLAB command 'expm' and by the fast matrix exponentiation method (Integrator 4.3.2) on a high dimensional example with $d_f = 100$. Accuracy-wise, the proposed method simulations yield results similar to VE (note that fast variables are not fully resolved due to a coarse time step that is larger than their periods). Speed-wise, Variational Euler, the proposed methods via 'expm' and via symplectic exponentiation respec-

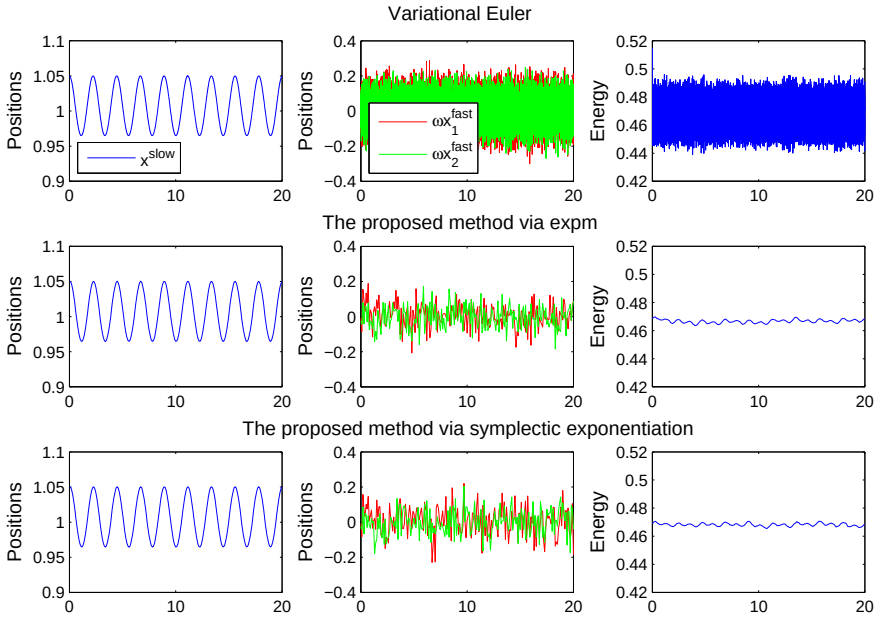


Figure 4.7: Simulations of a non-diagonal fast frequency high-dimensional example (4.87) by Variational Euler, the proposed method via MATLAB matrix exponentiation 'expm,' and the proposed method via fast matrix exponentiations ($n = 10$). Fast variable dimensionality is $d_f = 100$. $\omega = 1000$. VE uses $h = 0.1/\omega$ and the proposed method uses $H = 0.1$, $q(0) = 1.05$, $x(0)$ is a $d_f + 1$-dimensional vector with independent and identically distributed components that are normal random variables with zero mean and variance of $1/\omega/\sqrt{d_f}$ (so that energy is bounded), and initial momenta are zero. Only trajectories of the first two fast variables were drawn for clarity.

tively spent 136.7, 66.0 and 12.0 seconds on the same integration, while 65.7 and 11.7 seconds were spent on matrix exponentiation operations in the latter two. Notice that if Coppersmith-Winograd [75] is used to replace MATLAB matrix multiplication, the number 11.7 should be further reduced. In spite of that, the proposed method with the proposed matrix exponentiation scheme already holds a dominant speed advantage, and this advantage will be even more significant if ω and/or d_f is further increased (results not shown).

4.3.11 Additional details about the alternative matrix exponentiation scheme based on updating

We will present in Integrator 4.3.3 an alternative (symplectic) way for computing $F_{3,k}$ and $G_{2,k,i}$. This alternative is based on iteratively updating the matrix exponential from the computation at the previous step. We will first demonstrate its full version, and then provide a simple approximation which is not exactly symplectic on all variables but symplectic on the fast variables (in the sense of a symplectic submanifold) and exhibits satisfactory long time performance in numerical experiments.

Lemma 4.3.8. *Define:*

$$\begin{bmatrix} \alpha(t) & \beta(t) & \gamma(t) \\ 0 & F_2(t) & G_2(t) \\ 0 & 0 & F_3(t) \end{bmatrix} := \exp\left(\begin{bmatrix} -N^T & MJ & 0 \\ 0 & -N^T & M \\ 0 & 0 & N \end{bmatrix} t \right) \tag{4.88}$$

Then for any H, we have $-F_3(H)^T\gamma(H) = \int_0^H F_3^T(s)M(-JG_2(s))\,ds$.

Proof. Differentiating (4.88) with respect to t and equating each matrix component on left and right hand sides, we obtain:

$$\begin{cases} \dot{\alpha} = -N^T\alpha \\ \dot{F}_2 = -N^T F_2 \\ \dot{F}_3 = NF_3 \\ \dot{\beta} = -N^T\beta + MJF_2 \\ \dot{G}_2 = -N^T G_2 + MF_3 \\ \dot{\gamma} = -N^T\gamma + MJG_2 \end{cases} \tag{4.89}$$

where the initial conditions obviously are $\alpha(0) = I, F_2(0) = I, F_3(0) = I, \beta(0) = 0, G_2(0) = 0, \gamma(0) = 0$.

Solving these inhomogeneous linear equations leads to known results including

$F_2(t) = \exp(-N^T t)$, $F_3(t) = \exp(Nt)$ and $G_2(t) = \int_0^t \exp(-N^T(t-s))M\exp(Ns)\,ds$, as well as new results such as

$$\gamma(t) = \int_0^t \exp(-N^T(t-s))MJG_2(s)\,ds, \tag{4.90}$$

which is equivalent to

$$-F_3(H)^T\gamma(H) = \int_0^H F_3(s)^T M(-JG_2(s))\,ds \tag{4.91}$$

□

Lemma 4.3.9. *If $M = M^T$, $F_2^T F_3 = I$ and $\partial F_3 = -JG_2$, such as those derived from N and M defined in Section 4.3.4, then*

$$\begin{aligned}\partial G_2(H) = F_2(H)\Big(&-(F_3(H)^T\gamma(H))^T - F_3(H)^T\gamma(H) + \int_0^H F_3(s)^T\partial M F_3(s)\,ds \\ &-(-JG_2(H))^T G_2(H)\Big)\end{aligned} \tag{4.92}$$

Proof. By Leibniz's rule

$$\partial G_2(H) = [F_3(H)^T]^{-1}\left(\partial\left(F_3(H)^T G_2(H)\right) - \partial F_3(H)^T G_2(H)\right) \tag{4.93}$$

By the definition of F_3 and G_2, this is

$$\partial G_2(H) = F_2(H)\left(\partial\left(\int_0^H F_3(s)^T M F_3(s)\,ds\right) - (-JG_2(H))^T G_2(H)\right), \tag{4.94}$$

in which

$$\begin{aligned}&\partial\left(\int_0^H F_3(s)^T M F_3(s)\,ds\right) = \int_0^H \partial F_3(s)^T M F_3(s)\,ds + \int_0^H F_3(s)^T M\partial F_3(s)\,ds + \int_0^H F_3(s)^T\partial M F_3(s)\,ds \\ &= \int_0^H (-JG_2(s))^T M F_3(s)\,ds + \int_0^H F_3(s)^T M(-JG_2(s))\,ds + \int_0^H F_3(s)^T\partial M F_3(s)\,ds \\ &= -(F_3(H)^T\gamma(H))^T - F_3(H)^T\gamma(H) + \int_0^H F_3(s)^T\partial M F_3(s)\,ds\end{aligned} \tag{4.95}$$

for $\gamma(H)$ defined in Lemma 4.3.8. $\square$

Remark 4.3.11. $\int_0^H F_3(s)^T \partial M F_3(s)\, ds = \tilde{F}_3(H)^T \tilde{G}_2(H)$ *can be computed by again using the trick of:*

$$\begin{bmatrix} \tilde{F}_2(t) & \tilde{G}_2(t) \\ 0 & \tilde{F}_3(t) \end{bmatrix} := \exp\left(\begin{bmatrix} -N^T & \partial M \\ 0 & N \end{bmatrix} t \right) \tag{4.96}$$

Of course, to get $B'_{0,i,j} = \partial_j B_{0,i}$, *we use the fact that* $B_{0,i} = G_{2,0,i}(H/2^n)$.

Lemma 4.3.10. *Suppose* $q^{fast}_{(k+1)'}$, $p^{fast}_{(k+1)'}$, $q^{slow}_{(k+1)'}$, $p^{slow}_{(k+1)'}$ *are obtained from* $q^{fast}_{k'}$, $p^{fast}_{k'}$, $q^{slow}_{k'}$, $p^{slow}_{k'}$ *by Integrator 4.3.1 with* $F_{3,k}$ *and* $G_{2,k,i}$ *satisfying* $F_{3,k}^T J F_{3,k} = J$ *and* $G_{2,k,i} = -J \frac{\partial}{\partial q^{slow}_{k',i}} F_{3,k}$, *then*

$$\frac{\partial q^{slow}_{(k+1)',i}}{\partial q^{slow}_{k',j}} = I + \frac{H}{2} \begin{bmatrix} q^{fast}_{k'} \\ p^{fast}_{k'} \end{bmatrix} \left(G_{2,k,j}(H)^T J G_{2,k,i}(H) + F_{3,k}(H)^T \frac{\partial}{\partial q^{slow}_{k'}} G_{2,k}(H) \right) \begin{bmatrix} q^{fast}_{k'} \\ p^{fast}_{k'} \end{bmatrix} \tag{4.97}$$

Proof. Using chain rule, we have:

$$\frac{\partial q^{slow}_{(k+1)',i}}{\partial q^{slow}_{k',j}} = I + \frac{H}{2} \times \begin{bmatrix} q^{fast}_{k'} \\ p^{fast}_{k'} \end{bmatrix} \left(\frac{\partial}{\partial q^{slow}_{k',j}} F_{3,k}(H)^T G_{2,k,i}(H) + F_{3,k}(H)^T \frac{\partial}{\partial q^{slow}_{k',j}} G_{2,k,i}(H) \right) \begin{bmatrix} q^{fast}_{k'} \\ p^{fast}_{k'} \end{bmatrix} \tag{4.98}$$

This simplifies to (4.97) because $G_{2,k,i} = -J \frac{\partial}{\partial q^{slow}_{k',i}} F_{3,k}$ and $-J^T = J$. $\square$

Integrator 4.3.3. *Iterative matrix exponentiation scheme (alternative to Integrator 4.3.2) that obtains* $F_{3,k}$ *and* $G_{2,k,i}$ *via symplectic updates.* k *is the same index as the one used in Integrator 4.3.1.* $n \geq 1$ *is an integer controlling the accuracy of matrix exponential approximations.*

1. *At the beginning of simulation, let* $q^{slow}_{0'} = q^{slow}_0 + H p^{slow}_0$ *and evaluate* $K_0 := K(q^{slow}_{0'})$ *and* $\partial_i K_0 := \frac{\partial}{\partial q^{slow}_{0',i}} K(q^{slow}_{0'})$ *(*$i = 1, \ldots, d_s$*). Calculate*

$\begin{bmatrix} A_0 & B_{0,i} \\ 0 & C_0 \end{bmatrix} := \exp\left(\begin{bmatrix} -N_0^T & M_{0,i} \\ 0 & N_0 \end{bmatrix} H/2^n\right)$ *by any favorite matrix exponentiation method (e.g., by the symplectic method introduced in Section 4.3.4), where* $N_0 = \begin{bmatrix} 0 & I \\ -\epsilon^{-1}K_0 & 0 \end{bmatrix}$ *and* $M_{0,i} = \begin{bmatrix} \epsilon^{-1}\partial_i K_0 & 0 \\ 0 & 0 \end{bmatrix}$.

2. *Compute* $B'_{0,i,j} = \frac{\partial}{\partial q^{slow}_{0',j}} B_{0,i}$. *One cheap way to do so is to use Lemma 4.3.9 with Remark 4.3.11.*

3. *Start the updating loop, with the step count indicated by* k *starting from 1; let* $q_1^{slow,fast} = q_0^{slow,fast}$, $p_1^{slow,fast} = p_0^{slow,fast}$, *and* $\frac{q^{slow}_{1'}}{q^{slow}_{0'}} = I$;

4. *Carry out the* $q_k, p_k \mapsto q_{k'}, p_{k'}$ *half-step (in Integrator 4.3.1). Evaluate* $K_k := K(q^{slow}_{k'})$, *and let* $D_k := \begin{bmatrix} 0 & \epsilon^{-1}(K_k^T - K_{k-1}^T)H/2^n \\ 0 & 0 \end{bmatrix}$. *Define* $A_k := A_{k-1}\exp(D_k)$ *and use the equality* $\exp(D_k) = I + D_k$ *(since* D_k *is nilpotent); similarly, define* $C_k := C_{k-1}\exp(-D_k^T) = C_{k-1} - C_{k-1}D_k^T$;

5. *Let* $B_{k,i} = -J\frac{\partial C_k}{\partial q^{slow}_{k',i}}$, *which can be computed from known values using chain rule:*

$$\begin{aligned} B_{k,i} &= -J\frac{\partial(C_{k-1}(I+D_k))}{\partial q^{slow}_{k',i}} \\ &= -J\left(\sum_{j=1}^{d_s} \frac{\partial q^{slow}_{(k-1)',j}}{\partial q^{slow}_{k',i}} \frac{\partial C_{k-1}}{\partial q^{slow}_{k-1',j}}(I+D_k) + C_{k-1}\frac{\partial D_k}{\partial q^{slow}_{k',i}}\right) \\ &= \sum_{j=1}^{d_s} \frac{\partial q^{slow}_{(k-1)',j}}{\partial q^{slow}_{k',i}} B_{k-1,j}(I+D_k) + C_{k-1}\frac{\partial D_k}{\partial q^{slow}_{k',i}} \end{aligned} \tag{4.99}$$

To compute $\frac{\partial D_k}{\partial q^{slow}_{k',i}}$, *we need the derivatives of* K_k *and* K_{k-1} *with respect to* $q^{slow}_{k',i}$; *the former is trivial, and the latter again can be computed by chain rule:*

$$\frac{\partial K_{k-1}^T}{\partial q^{slow}_{k',i}} = \sum_{j=1}^{d_s} \frac{\partial q^{slow}_{(k-1)',j}}{\partial q^{slow}_{k',i}} \frac{\partial K_{k-1}^T}{\partial q^{slow}_{k-1',j}} \tag{4.100}$$

6. *$B'_{k,i,j}$ can be similarly computed from $B'_{k-1,i,j}$, $B_{k-1,i}$, C_{k-1} and D_k by repetitively applying chain rule. The detail is lengthy and hence omitted.*

7. *Let $F^1_{2,k} := A_k$, $G^1_{2,k,i} := B_{k,i}$, $F^1_{3,k} := C_k$, then repetitively apply $\begin{bmatrix} F^{j+1}_{2,k} & G^{j+1}_{2,k,i} \\ 0 & F^{j+1}_{3,k} \end{bmatrix} := \begin{bmatrix} F^{j}_{2,k} & G^{j}_{2,k,i} \\ 0 & F^{j}_{3,k} \end{bmatrix}^2 = \begin{bmatrix} F^j_{2,k}F^j_{2,k} & F^j_{2,k}G^j_{2,k,i} + G^j_{2,k,i}F^j_{3,k} \\ 0 & F^j_{3,k}F^j_{3,k} \end{bmatrix}$ for $j = 1, \ldots, n$, and finally define $F_{2,k} := F^{n+1}_{2,k}$, $G_{2,k,i} := G^{n+1}_{2,k,i}$, $F_{3,k} = F^{n+1}_{3,k}$.*

8. *Compute $\frac{\partial q^{slow}_{(k+1)',i}}{\partial q^{slow}_{k',j}}$ by using Lemma 4.3.10, so that it could be used by Step 5 for the next k. $\frac{\partial}{\partial q^{slow}_{k',j}} G_{2,k,i}(H)$ is computed based on the following:*

$$\frac{\partial}{\partial q^{slow}_{k',j}}(A_k B_{k,i} + B_{k,i} C_k) = -A_k B^T_{k,j} J A_k B_{k,i} + A_k B'_{k,i,j} + B'_{k,i,j} C_k - B_{k,i} J B_{k,j} \tag{4.101}$$

where the first term is due to $\frac{\partial A_k}{\partial q^{slow}_{k',j}} = -A_k B^T_{k,j} J A_k$, which is because $\partial A^T C + A^T \partial C = \partial(A^T C) = \partial I = 0$ and therefore $\partial A^T = -A^T \partial C C^{-1} = A^T J B C^{-1} = A^T J B A^T$. A similar trick of self multiplication applies to get the derivative of the 2^n-times product.

9. *Carry out the $q_{k'}, p_{k'} \mapsto q_{k+1}, p_{k+1}$ half-step update of numerical integration using $F_{2,k}$, $F_{3,k}$ and $G_{2,k,i}$, and then increase k by 1 and go to Step 4 until integration time is reached.*

$F_{3,k}$ and $G_{2,k,i}$ computed in this way (Integrator 4.3.3) will also satisfy Lemma 4.3.1 and render the integration symplectic on all variables. Proofs are omitted but they are analogous to those in Section 4.3.6, and all structures, such as reversibility, symplecticity of F_2 and F_3 (illustrated by corresponding lemmas), and the relation between F_3 and G_2, will be preserved as long as they are satisfied by $A_0, B_{0,i}, C_0$ (i.e., the initial matrix exponentiation is accurate).

In terms of efficiency, this method only uses one single matrix exponentiation operation and then keeps on updating it. Nevertheless, it is not easy to implement,

and its speed advantage is not dominant. However, if the requirement on symplecticity is not that strict and a small numerical error in the matrix exponential is allowed (recall an analogous case of the famous implicit mid-point integrator, in which implicit solves are in fact not done perfectly and continuously polluting the symplecticity), we could use the approximation of $\frac{\partial q^{slow}_{(k+1)',i}}{\partial q^{slow}_{k',j}} = I$. This will introduce a local error of $\mathcal{O}(Hn/2^n)$ in $G_{2,k,i}$ at each timestep (details omitted), but the local error in $F_{2,k}$ and $F_{3,k}$ is 0, and the method is symplectic on the submanifold of the fast variables (although not symplectic on all variables). The approximating method is:

Integrator 4.3.4. *An efficient approximation of Integrator 4.3.3:*

1. *At the beginning of simulation, let* $q^{slow}_{0'} = q^{slow}_0 + Hp^{slow}_0$ *and evaluate* $K_0 := K(q^{slow}_{0'})$ *and* $\partial_i K_0 := \partial_i K(q^{slow}_{0'})$ *(*$i = 1, \ldots, d_s$*) and calculate* $\begin{bmatrix} A_0 & B_{0,i} \\ 0 & C_0 \end{bmatrix} := \exp\left(\begin{bmatrix} -N_0^T & M_{0,i} \\ 0 & N_0 \end{bmatrix} H/2^n\right)$ *by any favorite matrix exponentiation method, where* $N_0 = \begin{bmatrix} 0 & I \\ -\epsilon^{-1}K_0 & 0 \end{bmatrix}$ *and* $M_{0,i} = \begin{bmatrix} \epsilon^{-1}\partial_i K_0 & 0 \\ 0 & 0 \end{bmatrix}$*; let* $q_1^{slow,fast} = q_0^{slow,fast}$ *and* $p_1^{slow,fast} = p_0^{slow,fast}$.

2. *Start the updating loop, with the step count indicated by k starting from 1;*

3. *Carry out the* $q_k, p_k \mapsto q_{k'}, p_{k'}$ *half-step. Evaluate* $K_k := K(q^{slow}_{k'})$ *and* $\partial_i K_k := \partial_i K(q^{slow}_{k'})$*, let* $D_k := \begin{bmatrix} 0 & \epsilon^{-1}(K_k^T - K_{k-1}^T)H/2^n \\ 0 & 0 \end{bmatrix}$ *and* $E_{k,i} := \begin{bmatrix} \epsilon^{-1}(\partial_i K_k - \partial_i K_{k-1})H/2^n & 0 \\ 0 & 0 \end{bmatrix}$*. Define* $\begin{bmatrix} A_k & B_{k,i} \\ 0 & C_k \end{bmatrix} := \begin{bmatrix} A_{k-1} & B_{k-1,i} \\ 0 & C_{k-1} \end{bmatrix} \times \exp\begin{bmatrix} D_k & E_{k,i} \\ 0 & -D_k^T \end{bmatrix}$ *and use the equality* $\exp\begin{bmatrix} D_k & E_{k,i} \\ 0 & -D_k^T \end{bmatrix} = \begin{bmatrix} I + D_k & E_{k,i} \\ 0 & I - D_k^T \end{bmatrix}$ *(because* $D_k E_{k,i} = 0$ *and* $E_{k,i} D_k^T = 0$*) to evaluate* $A_k = A_{k-1} + A_{k-1}D_k$*,* $B_{k,i} = B_{k-1,i} + A_{k-1}E_{k,i} - B_{k-1,i}D_k^T$*, and* $C_k = C_{k-1} - C_{k-1}D_k^T$*;*

4. *Let* $F^1_{2,k} := A_k$, $G^1_{2,k,i} := B_{k,i}$, $F^1_{3,k} := C_k$, *then repetitively apply* $\begin{bmatrix} F^{j+1}_{2,k} & G^{j+1}_{2,k,i} \\ 0 & F^{j+1}_{3,k} \end{bmatrix} := \begin{bmatrix} F^{j}_{2,k} & G^{j}_{2,k,i} \\ 0 & F^{j}_{3,k} \end{bmatrix}^2 = \begin{bmatrix} F^j_{2,k}F^j_{2,k} & F^j_{2,k}G^j_{2,k,i} + G^j_{2,k,i}F^j_{3,k} \\ 0 & F^j_{3,k}F^j_{3,k} \end{bmatrix}$ *for* $j = 1, \ldots, n$, *and finally define* $F_{2,k} := F^{n+1}_{2,k}$, $G_{2,k,i} := G^{n+1}_{2,k,i}$, $F_{3,k} = F^{n+1}_{3,k}$.

5. *Carry out the* $q_{k'}, p_{k'} \mapsto q_{k+1}, p_{k+1}$ *half-step update of numerical integration using* $F_{2,k}$, $F_{3,k}$ *and* $G_{2,k,i}$, *and then increase* k *by 1 and go to Step 3 until integration time is reached.*

Numerical experiments presented in Sections 4.3.8, 4.3.9 and 4.3.10 are repeated using this approximating integrator. Energy preservations are as good as before, and slow trajectories show no significant deviation, suggesting no significant effect of the approximated symplecticity (detailed results omitted). This approximation, on the other hand, allows a choice of an even smaller n, such as $n = 5$ for the previous examples, which results in a further speed-up.

Chapter 5

SyLiPN: Symplectic, linearly-implicit, and stable integrators, with applications to fast symplectic simulations of constrained dynamics

In the special case in which the system is stiff but admitting a trivial (Dirac-distributed) fast dynamics, implicit methods are enough for its coarse-step-integration. We propose a way to avoid expensive nonlinear solves in implicit methods, and yet keep the stability and symplecticity. This method applies to, for instance, the numerical integration of constrained dynamics, which could be modeled by a subclass of differential algebraic equations (DAEs).

Most results in this section can be found in a submitted manuscript [279].

5.1 Introduction

Implicit integrators are widely used for stable integrations, and the symplectic members of them have been successful in long time integrations of mechanical systems [128]. However, implicit methods require solving nonlinear systems, and therefore are generally much slower than linearly-implicit methods, because the latter only ask for solving linear systems, which can be carried out by various fast algorithms (see for instance [9] and references therein). For symplectic integrators, if one solves nonlinear equations partially (for instance, by carrying out only the first step in a gradient method [231]), or linearizes the nonlinearity (which is in fact similar to the former), symplecticity will in general be lost, resulting in an

unsatisfactory long time performance, such as drifted energy and momentum, etc.

To overcome these drawbacks, we propose a family of linearly-implicit and symplectic integrators that inherits the stability property of implicit methods. These methods are obtained via linearizing the push-forward of the Newmark family of implicit integrators [214, 257, 192]. We call them SyLiPN (Symplectic Linearized Push-forward Newmark).

As an important application of SyLiPN, we propose an efficient way to integrate mechanical systems with holonomic constraints. Since generalized coordinate approaches (e.g., [155]) and Lagrange multiplier methods (e.g., [237, 8, 138, 207]) are both implicit in general, we suggest a faster approach: first, model the constrained dynamics as stiff differential equations without algebraic constraints by replacing the rigid constraints by stiff springs (this model was proposed in [244] and could be dated back to the idea of penalty method, which was for instance reviewed in [230]); then, integrate the modified system by SyLiPN using a macroscopic timestep, which needs not to resolve the stiffness. Because the fast dynamics here (fast in the sense defined in Chapter 2) is merely a point distribution (as stiffness goes to infinity), the slow dynamics could be well captured by an implicit method [184], such as SyLiPN. Since a large timestep is used and each step is only linearly-implicit, constrained dynamics can be rapidly integrated in a symplectic way.

Two demonstrations, respectively on a double pendulum and a chain of many pendulums (an approximation to a continuous rope), are included. SyLiPN exhibits clear speed advantage. In addition, SyLiPN appears to be much more accurate than SHAKE when the integration timestep is very small. Similar applications, such as inexpensive large time-steppings of rods and shells with preserved momenta and nearly-preserved energy, could be useful in computer graphics and structure dynamics.

5.2 SyLiPN: Symplectic, linearly-implicit, and stable integrators

Consider the numerical integration of a mechanical system:

$$M\ddot{q} = -\nabla V(q) \tag{5.1}$$

where $q \in Q$ is the configuration and $V \in \mathcal{C}^2(Q)$ is the potential energy function. Oftentimes $Q = \mathbb{R}^n$, and in this case the mass M is indicated by a n-by-n matrix.

The following Newmark family of algorithms have been widely used in structure dynamics [214]:

Integrator 5.2.1. ***Newmark:***

$$\begin{cases} q_{k+1} &= q_k + h\dot{q}_k + \frac{h^2}{2}[(1-2\beta)a_k + 2\beta a_{k+1}] \\ \dot{q}_{k+1} &= \dot{q}_k + h[(1-\gamma)a_k + \gamma a_{k+1}] \\ a_k &= M^{-1}(-\nabla V(q_k)) \end{cases} \tag{5.2}$$

It was known [192] that Newmark is 2nd-order accurate when $\gamma = 1/2$ and 1st-order otherwise, and it is generally implicit when $\beta \neq 0$. It was also shown [161] that Newmark is variational for arbitrary β when $\gamma = 1/2$ (we will restrict ourselves to this case throughout this chapter). However, it is worth noticing that the symplectic form that Integrator 5.2.1 preserves is not the canonical one. In fact, it was shown [257, 192] that if one pushes forward the Newmark integrator by the map $\eta : TQ \to TQ$:

$$\eta(q, v) = (q + \beta h^2 M^{-1} \nabla V(q), v) \quad , \tag{5.3}$$

then we obtain an integrator that preserves the canonical symplectic form on T^*Q:

Integrator 5.2.2. ***Push-forward Newmark:***

$$\begin{cases} x_{k+1} &= x_k + hv_k + \frac{1}{2}h^2 a_k \\ v_{k+1} &= v_k + \frac{1}{2}h(a_k + a_{k+1}) \\ a_k &= -M^{-1}\nabla V(x_k + \beta h^2 a_k) \end{cases} \tag{5.4}$$

The Newmark and push-forward Newmark schemes can be shown to be unconditionally linearly stable if $\beta \geq 1/4$ [64, 257], and Newmark was further shown to be nonlinearly stable in the same case under several assumptions [149]. In addition, push-forward Newmark was shown to be stable near stable fixed points in general nonlinear settings unless specific resonances occur [256]. Nevertheless, there are nonlinear cases in which Newmark is no longer stable [98, 172]. In fact, few convergent methods are unconditionally stable for arbitrary nonlinear systems to the authors' knowledge (e.g., see a discussion in [298]).

Now, linearize the nonlinear force in push-forward Newmark at each step by Taylor expansion at x_k, so that a nonlinear implicit equation becomes a linear one. One obtains:

Integrator 5.2.3. ***Symplectic Linearized Push-forward Newmark (SyLiPN):***

$$\begin{cases} x_{k+1} &= x_k + hv_k + \frac{1}{2}h^2 a_k \\ v_{k+1} &= v_k + \frac{1}{2}h(a_k + a_{k+1}) \\ a_k &= -M^{-1}\nabla V(x_k) - M^{-1}HessV(x_k)\beta h^2 a_k \end{cases} \tag{5.5}$$

Remark 5.2.1. *Notice that the third line, i.e., the force evaluation, could be rewritten as*

$$a_k = -(I + M^{-1}HessV(x_k)\beta h^2)^{-1} M^{-1}\nabla V(x_k) \tag{5.6}$$

This evaluation, however, should be executed by solving a symmetric linear system instead of inverting a matrix due to the consideration of computational efficiency. In this sense, SyLiPN is linearly implicit.

Since Newmark or Push-forward Newmark requires solving a nonlinear system at each step, SyLiPN exhibits a speed advantage. We will quantify this advantage numerically in Section 5.4.

Theorem 5.2.1. *SyLiPN (Integrator 5.2.3) is linearly unconditionally stable if* $\beta \geq 1/4$.

Proof. This is straightforward, because for linear test problems in which $V(\cdot)$ is quadratic, SyLiPN is identical to Push-forward Newmark, which is equivalent to Newmark in terms of stability. □

Remark 5.2.2. *[149] uses an energy bound based on an assumption to demonstrate the nonlinear stability of Newmark. The same energy bound applies to SyLiPN, because Newmark and SyLiPN differ in force estimations by only a high-order term of* $O(h^2)$*, which will not affect the leading term of the energy bound. The assumption introduced there, however, can not be checked a priori for either Newmark or SyLiPN. Possible violations of this assumption may result in a nonlinear instability of Newmark or SyLiPN. As commented before, unconditional stability for arbitrary nonlinear systems is beyond the scope of current research.*

Remark 5.2.3. *For possible improvements of nonlinear stability, one may resort to linearizations of more stable methods (such as those in [171]). However, few methods after the linearization are still symplectic.*

Theorem 5.2.2. *If* $V \in \mathcal{C}^3(Q)$*, SyLiPN (Integrator 5.2.3) has a 3rd-order local error, i.e., start with* x_k, v_k *at time* kh *and denote by* $\tilde{x}_{k+1}, \tilde{v}_{k+1}$ *the exact solution at time* $(k+1)h$ *and by* x_{k+1}, v_{k+1} *the numerical solution after one-step update, then* $\tilde{x}_{k+1} - x_{k+1} = O(h^3)$ *and* $\tilde{v}_{k+1} - v_{k+1} = O(h^3)$.

Proof. Assume without loss of generality that $M = I$. Writing $a(\cdot) = -\nabla V(\cdot)$, we have

$$a_k = a(x_k)/(1 - a'(x_k)\beta h^2) = a(x_k) + O(h^2) \tag{5.7}$$

and

$$a(x_{k+1}) = a(x_k)+(x_{k+1}-x_k)a'(x_k)+O\left((x_{k+1}-x_k)^2\right) = a(x_k)+hv_ka'(x_k)+O(h^2) \tag{5.8}$$

Since the exact dynamics is governed by

$$\begin{cases} \dot{x} = v \\ \dot{v} = a(x) \end{cases}, \tag{5.9}$$

if smoothness of the solution is assumed, we obtain by Taylor expansion that

$$\begin{aligned} \tilde{x}_{k+1} &= x_k + hv_k + \frac{h^2}{2}\dot{v_k} + O(h^3) \\ &= x_k + hv_k + \frac{h^2}{2}\left(a_k + O(h^2)\right) + O(h^3) \\ &= x_k + hv_k + \frac{h^2}{2}a_k + O(h^3) = x_{k+1} + O(h^3) \end{aligned} \tag{5.10}$$

$$\begin{aligned} \tilde{v}_{k+1} &= v_k + ha(x_k) + \frac{h^2}{2}a'(x_k)v_k + O(h^3) \\ &= v_k + \frac{h}{2}(a(x_k) + a(x_k) + ha'(x_k)v_k) + O(h^3) \\ &= v_k + \frac{h}{2}(a_k + O(h^2) + a(x_{k+1}) + O(h^2)) + O(h^3) \\ &= v_k + \frac{h}{2}(a_k + O(h^2) + a_{k+1} + O(h^2)) + O(h^3) = v_{k+1} + O(h^3) \end{aligned} \tag{5.11}$$

□

Remark 5.2.4. *By the famous Lax-Richtmyer equivalence theorem [174], an $\mathcal{O}(h^p)$ local error (also known as 'consistency' when $p \geq 2$) together with stability will lead to an $\mathcal{O}(h^{p-1})$ global error (i.e., 'convergence' for $p \geq 2$). For our case, if SyLiPN is stable, then it is 2nd-order convergent (with an $\mathcal{O}(h^2)$ global error).*

Theorem 5.2.3. *SyLiPN (Integrator 5.2.3) is symplectic.*

Proof. The Jacobian of the one-step update given by (5.5) can be computed. A

lengthy calculation shows that (details are omitted here):

$$\begin{bmatrix} \frac{\partial x_{k+1}}{\partial x_k} & \frac{\partial x_{k+1}}{\partial y_k} \\ \frac{\partial y_{k+1}}{\partial x_k} & \frac{\partial y_{k+1}}{\partial y_k} \end{bmatrix}^T J \begin{bmatrix} \frac{\partial x_{k+1}}{\partial x_k} & \frac{\partial x_{k+1}}{\partial y_k} \\ \frac{\partial y_{k+1}}{\partial x_k} & \frac{\partial y_{k+1}}{\partial y_k} \end{bmatrix} = J \tag{5.12}$$

where $J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is the canonical symplectic matrix. This proves that the integrator is symplectic. □

Remark 5.2.5. *Linearizing Newmark or implicit midpoint will not result in a symplectic method, although the resulting integrator will be linearly-implicit, stable and 2nd-order. It is rare that the linearization of an implicit method is symplectic.*

5.3 Linearly-implicit symplectic simulation of constrained dynamics

Now consider again a mechanical system in which particles with a mass matrix M positioned at q are evolving on a potential landscape $V(\cdot)$, but this time a (possibly vectorial) holonomic constraint $g(q) = 0$ is present. One mathematical way to represent this problem is via Hamilton's principle on a constrained manifold: define the action functional:

$$\mathcal{S}(q(t)) := \int_a^b \frac{1}{2}\dot{q}(t)^T M \dot{q}(t) - V(q(t))\, dt \tag{5.13}$$

and look for critical trajectory on the constrained manifold, i.e., solve the equation:

$$\delta\mathcal{S}/\delta q = 0 \text{ for } q(t) \in g^{-1}(0)\ \forall t \tag{5.14}$$

Traditional approaches to simulate such a system include: introducing generalized coordinates (e.g., [155]), so that the constraints completely disappear; using Lagrange multipliers (e.g., SHAKE [237], RATTLE [8], SETTLE [138], LINCS [207]), so that the problem converts to the numerical simulation of a DAE system;

and so on. These methods allow an $o(1)$ integration step, but in general both involve solving nonlinear systems that slows down the computation.

We relax the rigid constraint by instead using a stiff spring to reinforce the constraint up to a small deviation (e.g., [244] and [230]). More precisely, modify the potential energy $V(q)$ to be $V(q) + \frac{1}{2}\omega^2 g(q)^2$, and then simulate the modified mechanical system without the constraint:

$$M\ddot{q} = -\nabla V(q) - \omega^2 g(q)\nabla g(q) \tag{5.15}$$

The idea is that, as $\omega \to \infty$, trajectories of this modified system will approximate those of the constrained dynamics. Due to energy conservation in the modified system, $\dot{q}M\dot{q}/2 + V(q) + \frac{1}{2}\omega^2 g(q)^2$ has a constant bounded value due to the initial condition, and therefore $\frac{1}{2}\omega^2 g(q)^2 = \mathcal{O}(1)$ unless instability happens. As a consequence, $g(q) = \mathcal{O}(1/\omega)$ and the constraint will be satisfied approximately. In the sense of separation of timescales (formally defined in Chapter 2), small oscillations around the constrained values are of frequency $\mathcal{O}(1/\omega)$ and correspond to a fast process, which converge to a Dirac point distribution (fixed value), whereas the slow process approximates the dynamics on the constrained manifold.

The link between this formalism and the approach of Lagrange multiplier can be understood as follows: assume the existence of a limiting solution $q(t)$ as $\omega \to \infty$, and then let $\lambda = \lim_{\omega\to\infty} -\omega^2 g(q)$, we obtain the following DAE system:

$$\begin{cases} M\ddot{q} = -\nabla V(q) + \lambda\nabla g(q) \\ g(q) = 0 \end{cases} \tag{5.16}$$

Due to the uniqueness of the solution, this asymptotic solution is identical to the continuous Lagrange multiplier solution, because they satisfy the same equation.

Note that the equivalence between the approaches of generalized coordinate and continuous Lagrange multiplier was established in [295]. Consequently, our model is formally equivalent to the generalized coordinate approach in non-pathological cases, and is therefore justified.

On a more technical note regarding non-pathological cases, one of the known necessary conditions for the existence of a limiting solution is that Takens chaos [271] does not happen. Since Takens chaos happens when multiple eigenfrequencies of the fast process are identical, this source of inaccuracy could be easily avoided by choosing different large values of ω for multiple constraints, i.e., to use a modified potential energy of $V(q) + \frac{1}{2}g(q)^T \mathrm{diag}(\omega_1, \omega_2, \ldots, \omega_n)g(q)$, where n is the number of constraints and ω_i's have distinct values. We did not observe any anomaly in our numerical experiments, and the trajectories of the modified system always approximated the constrained dynamics well when ω was large.

Numerically, one could use a textbook symplectic integrator, such as symplectic Euler (also known as Variational Euler or leapfrog) or Velocity-Verlet, together with a timestep of length of $o(1/\omega)$ to simulate the modified system. However, this is not optimally efficient for obvious reasons (although sometimes it is already faster than generalized coordinate or Lagrange multiplier approaches).

Alternatively, the above SyLiPN (Integrator 5.2.3) allows a large step of size $o(1)$ for a linearly-implicit integration, which naturally will be much faster when ω is large. The reason that SyLiPN with $o(1)$ timestep works for (5.15) is because the stiffness in this system results in a fast dynamics that converges to a point distribution (a precise definition of the sense of convergence can be found in Chapter 2), whose contribution to the slow dynamics therefore could be captured by an implicit method [184].

5.4 Numerical examples

5.4.1 Double pendulum

Implementation: Consider a double pendulum system. One way to represent the system is to use 4 degrees of freedom and 2 nonlinear constraints in Euclidian

coordinates. Writing in the above notations, we have

$$M = \begin{bmatrix} m_1 & 0 & 0 & 0 \\ 0 & m_1 & 0 & 0 \\ 0 & 0 & m_2 & 0 \\ 0 & 0 & 0 & m_2 \end{bmatrix} \tag{5.17}$$

$$V(x_1, y_1, x_2, y_2) = -gy_1 - gy_2 \tag{5.18}$$

$$g(x_1, y_1, x_2, y_2) = \begin{bmatrix} x_1^2 + y_1^2 - L_1^2 \\ (x_2 - x_1)^2 + (y_2 - y_1)^2 - L_2^2 \end{bmatrix} \tag{5.19}$$

where m_1, m_2 are two masses, g is the gravitational constant, and L_1, L_2 are lengths of the two pendulums. To simplify our notations, we adopt a unitless convention and assume $m_1 = m_2 = g = 1$.

Implementation of our constraint-free approach on this system is straightforward (Eq. 5.15).

To integrate in generalized coordinates θ, ϕ, let $x_1 = L_1 \sin\theta, y_1 = -L_1 \cos\theta, x_2 = L_1 \sin\theta + L_2 \sin\phi, y_2 = -L_1 \cos\theta - L_2 \cos\phi$. Then the Lagrangian $L(q, \dot{q}) = \frac{1}{2}\dot{q}M\dot{q}^T - V(q)$ with $q = [x_1\ y_1\ x_2\ y_2]$ turns out to be

$$\tilde{L}(\theta, \phi, \dot{\theta}, \dot{\phi}) = \frac{1}{2}\big(2L_1^2\dot{\theta}^2 + L_2^2\dot{\phi}^2 + 2L_1L_2(\cos\theta\cos\phi + \sin\theta\sin\phi)\dot{\theta}\dot{\phi}\big) + 2L_1\cos\theta + L_2\cos\phi \tag{5.20}$$

in which the length constraints are intrinsically handled. Corresponding Euler-Lagrangian equations will give the constrained dynamics. Numerically, one approximates the action $\int_{kh}^{(k+1)h} \tilde{L}(\theta, \phi, \dot{\theta}, \dot{\phi})dt$ using a quadrature rule and obtains a discrete Lagrangian $\tilde{L}_d(\theta_k, \phi_k, \theta_{k+1}, \phi_{k+1})$. Applying the least action principle again, a set of discrete Euler-Lagrangian equations [192] are obtained. Notice that the mass matrix in the generalized coordinates is

$$\tilde{M}(\theta, \phi) = \begin{bmatrix} 2L_1^2 & L_1L_2(\cos\theta\cos\phi + \sin\theta\sin\phi) \\ L_1L_2(\cos\theta\cos\phi + \sin\theta\sin\phi) & L_2^2 \end{bmatrix}, \tag{5.21}$$

which is no longer constant but position dependent. As a consequence, variational

integrators given by discrete Euler-Lagrangian equations, even symplectic Euler or Verlet, will be implicit.

Regarding numerical approximations to the continuous Lagrange multiplier system (5.16), the well-known algorithms of SHAKE and RATTLE can be viewed as variational integrators with 1st-order and 2nd-order quadrature discretizations of the Lagrange-d'Alembert principle $\int \big(L(x_1, y_1, x_2, y_2) + \lambda \cdot g(x_1, y_1, x_2, y_2)\big)dt$, in which constraints are realized via a vectorial Lagrange multiplier $\lambda(t)$ [192]. These methods also involve an implicit solve at each step to compute the virtual force (λ) that reinforces the constraints.

Results: Figure 5.1 provides a comparison of different integration methods. Our constraint-free model (Subplots 1 and 3) produced results almost identical to SHAKE simulations, and they are also in good agreement with the benchmark produced by the generalized coordinate approach. Notice that a non-zero measured set of initial conditions leads to chaotic behaviors in this system [233]; therefore, a symplectic integration of the system is desired [60, 200]. On a similar note, an accurate integration is highly nontrivial even when the initial condition (e.g., the one here) is not in the chaotic region, as numerical errors may easily lead to chaotic regions from regular trajectories. None of the methods tested here exhibits such an errant behavior.

Speed-wise, Variational Euler (VE) on the constraint-free modified system, SHAKE, SyLiPN, and two generalized coordinate implicit VEs respectively took 34.1, 23.8, 1.0, 36.7 and 373.5 seconds for the above simulation (on a 2.4 GHz laptop running MATLAB 7.7 and 'fsolve' as the nonlinear solver). SHAKE, SyLiPN and generalized coordinate implicit VEs are able to use a large timestep independent of ω, whereas VE on the constraint-free modified system uses a small timestep to resolve the stiffness. SHAKE and generalized coordinate approach are based on solving nonlinear equations, which significantly solved down the computation. SyLiPN uses a large step and only involves solving linear systems, and therefore is superior in terms of computational efficiency.

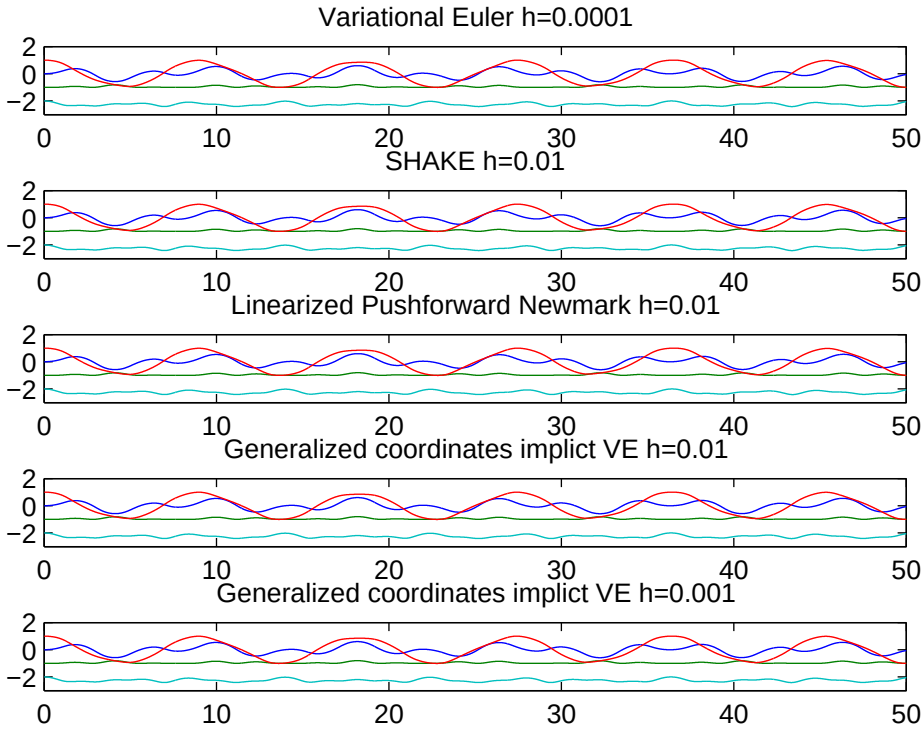


Figure 5.1: Comparison of $x_1(t), y_1(t), x_2(t), y_2(t)$ integrated by small step explicit Variational Euler on the modified system (Eq. 5.15), big step implicit SHAKE, big step linearly-implicit SyLiPN on the modified system, and big step and medium step (benchmark) Variational Eulers in generalized coordinates. Initial positions are $x_1(0) = 0, y_1(0) = -1, x_2(0) = 1, y_2(0) = -2$ and initial momenta are zero, $L_1 = 1$ and $L_2 = \sqrt{2}$, and total simulation time is 50. The modified system (corresponding to Subplots 1 and 3) uses $\omega = 1000$ although $\omega = 100$ produces no visible difference. SyLiPN uses $\beta = 0.4$. Subplot 1 uses $h = 0.1/\omega$ for stability, Subplot 2 uses $h = 0.01$ for stability too, Subplot 3 and 4 use $h = 0.01$ to match SHAKE, and Subplot 5 uses $h = 0.001$, a step much smaller than stability requirement in order to reduce numerical error and serve as a benchmark.

Two points regarding solving linear systems for a chain of pendulums are worth mentioning. First, the increase in computational cost when extending from two pendulums to finitely many will only be linear, which is the best one can expect. This is because the coefficient matrix in the linear system only has nonzero entries around its diagonal, due to the fact that both the original potential and the constraining potential involve only local interactions. Second, one could actually analytically pre-compute a position-dependent nonlinear function, which corresponds to the inversion of the matrix $I + \text{Hess}V(x)\beta h^2$, so that SyLiPN (Integrator 5.2.3) becomes entirely explicit. Doing so, however, will not result in a gain in speed, because the linear solve is less expensive than a force computation by evaluating the pre-computed matrix inversion and multiplying it by a vector. In fact, we counted the time elapses of both linear solves and explicit multiplications with the pre-computed inverse matrix in our numerical experiments, and they respectively take 1.0 and 14.0 seconds (the inverse matrix was generated automatically by Mathematica function 'Inverse', and therefore the code for computing it may not be optimized). In this sense, a linearly-implicit method is good enough.

5.4.2 Convergence test

By Theorem 5.2.2 and Remark 5.2.4, we know that SyLiPN is a 2nd-order integrator. However, this is only a quantification of its integration ability at small h. In this section, we will numerically investigate two additional questions:

(i) How does the integration error of SyLiPN scale with the timestep h when h is large?

(ii) Besides the error of ODE integration, the constrained dynamics model (5.15) is also just an approximation. Combining both approximations, how does the error of SyLiPN simulation of the constrained dynamics depend on h?

In addition, we found in our numerical experiments that SHAKE yields a significant error when h becomes sufficiently small, and this is due to numerical errors in solving nonlinear equations. On the other hand, SyLiPN does not exhibit this drawback.

Double pendulum: First, continue our numerical experiments on the double pendulum system (Section 5.4.1). The same initial condition and β will be used, but we decrease ω from 1000 to 100 to demonstrate the error in the constrained dynamics model, and we also decrease the simulation time from 50 to 10 so that the following investigation can be done within a short time. In addition, we further decrease h from 0.001 to 0.0001 in the generalized coordinate implicit VE so that the benchmark will be more accurate.

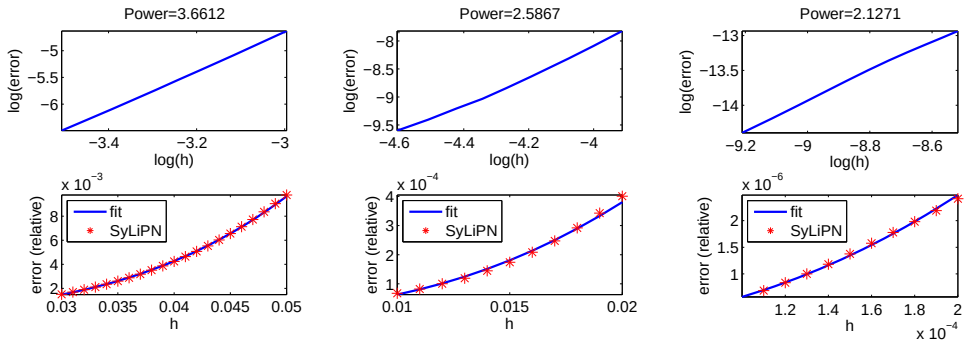
To study question (i), we compare SyLiPN integration of (5.15) with timestep h and a benchmark integration of (5.15) obtained by Variational Euler with a very small timestep $0.0001/\omega$. Values of h are enumerated, and differences between integrated positions at a fixed time ($t = 10$ in our case) are collected (measured in 2-norm and normalized by the 2-norm of the benchmark position). These pairs will approximate SyLiPN's integration error as a function of h. Assuming

$$\text{error} = C_1 h^{C_2} \tag{5.22}$$

for some constants $C_1 > 0$ and $C_2 > 0$, we fit a linear model $\log(\text{error}) = \log C_1 + C_2 \log h$ to obtain C_1 and C_2. C_2 is the power that we are interested in; for instance, a value of ≈ 2 means a 2nd-order global error.

As can be seen in Figure 5.2, when $h \gg h_0/\omega$ (where h_0/ω is the stability limit of Variational Euler for (5.15)), the convergence towards the solution to (5.15) is in fact faster than quadratic, and the bigger h is the faster the convergence. Of course, when $h < h_0/\omega$, the convergence rate converges to $C_2 \approx 2$, which is consistent with Theorem 5.2.2. Put together, the convergence rate decreases from a large value at big h to 2 at small h. This provides another piece of evidence explaining why SyLiPN with a large h is accurate.

To investigate question (ii), we carry out a similar analysis, but this time the benchmark is obtained by generalized coordinate implicit Variational Euler with a very small step $h = 0.0001$. As can be seen in Figure 5.3, when $h \gg h_0/\omega$, the convergence towards the solution to (5.14) is at least quadratic, and the smaller

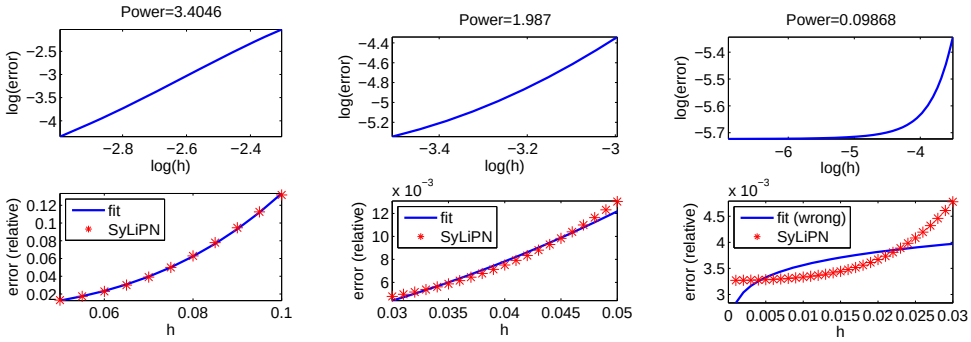


(a) *Large* $h = 0.03, 0.031, \ldots, 0.05 \gg h_0/\omega$ **(b) *Medium*** $h = 0.01, 0.011, \ldots, 0.02 \gg h_0/\omega$ **(c) *Very small*** $h = 0.0001, 0.00011, \ldots, 0.0002 < h_0/\omega$

Figure 5.2: Relative error of SyLiPN integration of (5.15) with (5.19) as a function of h. The straighter the line is in the log-log plot (first row), the better the error function fits in a power law (5.22).

the h, the closer the convergence rate approaches quadratic. Interesting behavior happens at $h \sim h_0/\omega$, where the error reaches a plateau and does not further decrease with h to 0. This is because the constrained dynamics model (5.15) introduces an $\mathcal{O}(1/\omega)$ error from (5.14). This error could be reduced by increasing ω (plot not shown), but an $\omega > 100$ (and a corresponding small h) is rarely necessary, unless a high precision beyond 1/1000 is in demand.

Moreover, we see in Figure 5.4 that a numerical implementation of SHAKE yields significant error when h is further decreased, while SyLiPN is not convergent to the benchmark either but its error is much smaller. In particular, in this experiment SHAKE yields $\sim \exp(-6)$ error when h is small, but when h further becomes smaller, the error grows to ~ 1 (i.e., 100%) at some point, and then decreases to $\sim \exp(-2)$ again and stablizes. This is due to the inevitable numerical errors in solving nonlinear systems in SHAKE: we plotted λ the Lagrange multiplier in SHAKE as a function of time (results not shown) and found that λ was 0 in the first few steps but then suddenly became very large; this is because h is so small that in the first few steps it appears to the nonlinear solver that the constraints are satisfied without any virtue force, but then the deviation of the



(a) ***Large*** $h = 0.05, 0.055, \ldots, 0.1 \gg h_0/\omega$

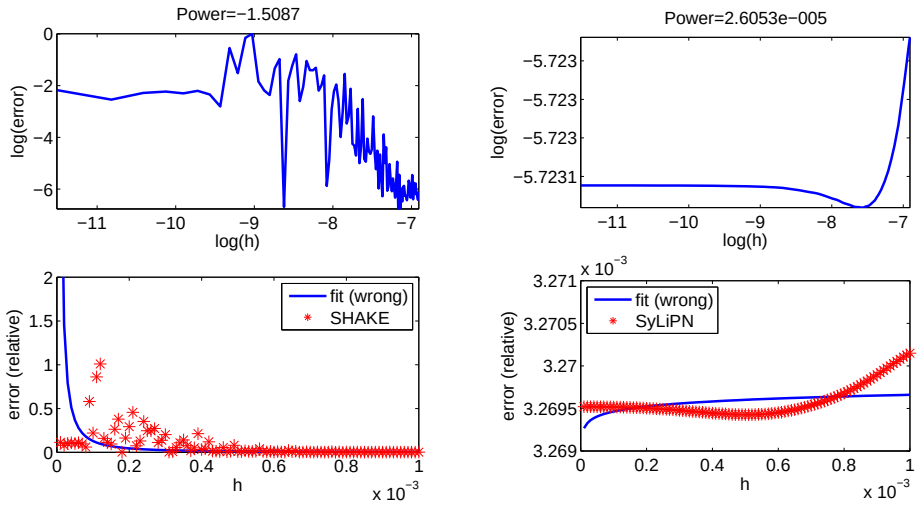
(b) ***Medium*** $h = 0.03, 0.031, \ldots, 0.05 \gg h_0/\omega$

(c) ***Small*** $h = 0.001, 0.002, \ldots, 0.03 \gtrsim h_0/\omega$. ***The fit doesn't work because the error does not go to 0 as*** $h \to 0$ ***(and hence the model*** (5.22) ***is wrong).***

Figure 5.3: Relative error of SyLiPN simulation of the constrained dynamics (5.14) with (5.19) as a function of h. The straighter the line is in the log-log plot (first row), the better the error function fits in a power law (5.22).

constraint from being satisfied accumulates, and this eventually leads to a sudden overshot of the Lagrange multiplier. On the contrary, SyLiPN does not exhibit such a problem, and its error converges to a small non-zero value because the constraint is only satisfied approximately with an error of $\mathcal{O}(1/\omega)$, and this error can be reduced by choosing a large ω.

Uniform circular motion: To rule out the possibility of an inaccurate benchmark, we repeated the above experiments on a particle with no potential energy but constrained by $x^2 + y^2 = 1$, whose dynamics can be solved exactly and corresponds to a uniform circular motion. We obtained similar results (with slightly different numerical values of powers), and SHAKE still exhibits inaccuracy at tiny h, but we will not repeat the details.



(a) *SHAKE* **(b) *SyLiPN with constrained dynamics model*** (5.15)

Figure 5.4: Relative error of SHAKE and SyLiPN simulation of the constrained dynamics (5.14) with (5.19) as a function of h. h is chosen from $0.00001 \ll h_0/\omega$ to $0.001 \approx h_0/\omega$ with an increment of 0.00001. The benchmark is obtained by generalized coordinate implicit Variational Euler with a very small step $h = 0.0001$.

5.4.3 High-dimensional case: A chain of many pendulums

Now consider a high-dimensional generalization: a chain of finitely many pendulums (which approximates a rope, except for that ropes in reality are subject to dissipations and therefore not chaotic). The system is similarly modeled by (5.15)

with:

$$M = \begin{bmatrix} m_1 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & m_1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots & \vdots \\ \vdots & \vdots & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & m_n & 0 \\ 0 & 0 & \cdots & \cdots & 0 & m_n \end{bmatrix} \tag{5.23}$$

$$V(x_1, y_1, \cdots, x_n, y_n) = -\sum_{i=1}^{n} g y_i \tag{5.24}$$

$$g(x_1, y_1, \cdots, x_n, y_n) = \begin{bmatrix} x_1^2 + y_1^2 - L_1^2 \\ (x_2 - x_1)^2 + (y_2 - y_1)^2 - L_2^2 \\ \vdots \\ (x_n - x_{n-1})^2 + (y_n - y_{n-1})^2 - L_n^2 \end{bmatrix} \tag{5.25}$$

where n indicates the total number of pendulums. We again assume without loss of generality that $m_i = 1$ and $g = 1$.

Figure 5.5 provides a comparison between SHAKE and SyLiPN. The system is chaotic, meaning that even the same converging integrator with slightly different time step lengths will eventually produce completely different trajectories, and therefore we terminate the integration before chaotic behavior starts to manifest so that the comparison still makes sense. Such a termination time is decided so that SHAKE with different integration step lengths produces the same trajectory, but simulations beyond that time will yield significant deviations. SyLiPN agrees well with SHAKE till this termination time. An animation that compares the simulations is available at `http://www.youtube.com/watch?v=naTStCPuW9M`.

Speed-wise, SHAKE with $h = 0.1$, $h = 0.05$, $h = 0.2$ and SyLiPN with $h = 0.1$ respectively spent 18.9, 37.1, 11.5 and 1.5 seconds on the above simulation (on a 2.4 GHz laptop running MATLAB 7.7 and 'fsolve' as the nonlinear solver). Again, SyLiPN based on linear solves demonstrates a clear speed advantage.

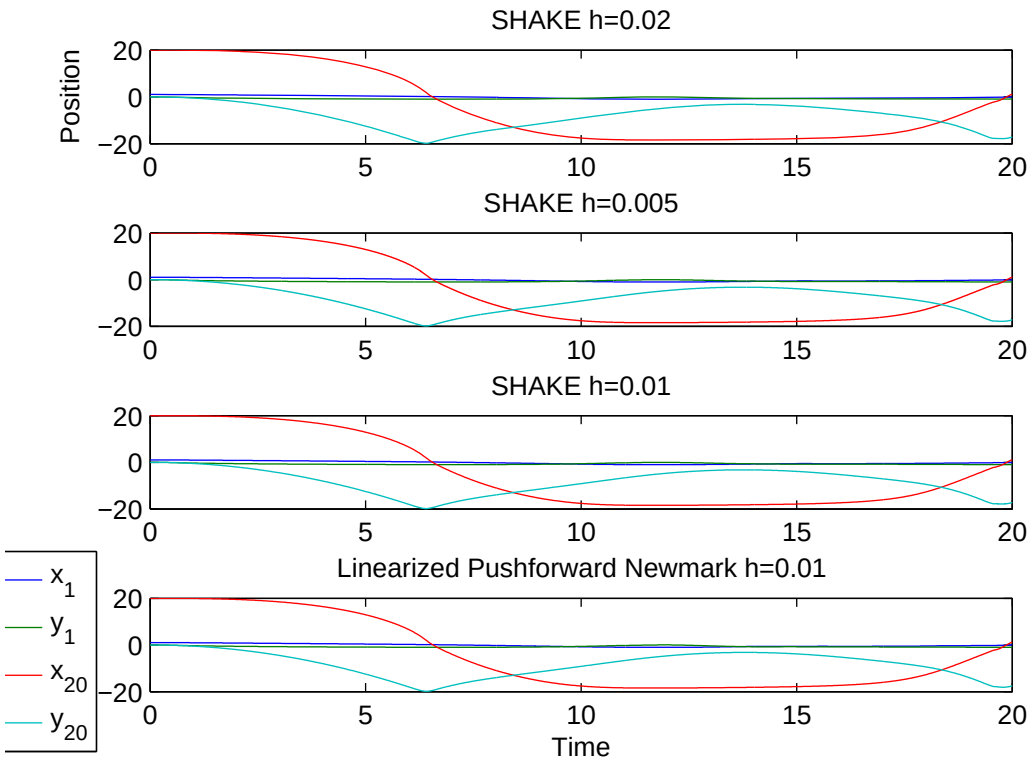


Figure 5.5: Comparison of $x_1(t), y_1(t), x_n(t), y_n(t)$ integrated by implicit SHAKE with $h = 0.05$, $h = 0.1$ and $h = 0.2$ and linearly-implicit SyLiPN with $h = 0.1$ on the modified system (5.15). $n = 20$; initial positions are $x_i(0) = i, y_i(0) = 0$ for $i = 1, 2, \ldots, n$ and initial momenta are zero; $L_i = 1$; the modified system uses $\omega = 100$; total simulation time is 20; SyLiPN uses $\beta = 0.4$.

Chapter 6

Temperature and friction accelerated sampling

In a stochastic setting, accurately obtaining the statistics of an ensemble of numerical solutions is often desired. This accuracy is, of course, distinct from a trajectory-wise accuracy. In fact, it is not only possible but also computationally beneficial to ignore the correctness of each single trajectory but still obtain a statistically correct ensemble. For the purpose of sampling from a statistical distribution, this is perfectly fine, because, after all, the dynamics is artificially introduced only in order to enable sampling.

Based on this philosophy, we propose a method to accelerate the convergence of a Langevin process towards its corresponding Boltzmann-Gibbs (B-G) distribution, so that an efficient B-G sampling could be achieved.

Most results in this chapter are included in a submitted manuscript [277].

6.1 Introduction

Specifically, consider the following Langevin Stochastic Differential Equations

$$\begin{cases} dq &= pdt \\ dp &= -\nabla V(q)dt - cpdt + \sqrt{2c/\beta}dW \end{cases} \tag{6.1}$$

where $p, q \in \mathbb{R}^d$ represent position and momentum, $V(\cdot)$ is the potential energy, W is a standard Wiener process, M is the mass matrix, c is a positive semi-definite $d \times d$ matrix indicating the damping coefficient, $\beta \in \mathbb{R}^+$ is the inverse of temperature, and c and β are constants that do not depend on q or p.

It is known that the stochastic process defined by (6.1) has an invariant distribution of Boltzmann-Gibbs defined by:

$$d\mu = Z^{-1} \exp(-\beta H(q,p)) dq dp \tag{6.2}$$

where $Z = \int_{T^*\mathbb{R}^d} \exp(-\beta H(q,p)) dq dp$ is the partition function, and $H(q,p) = p^T M^{-1} p/2 + V(q)$ is the Hamiltonian function.

When the solution of (6.1) is also geometrically ergodic with respect to μ (we refer to [197] and [198] for sufficient conditions on the potential V), it is then natural to use long-time trajectories of (6.1) as approximate samples of B-G distribution.

One important thing to notice is that being able to sample from B-G enables sampling an **arbitrary** smooth-enough probability density function. The trick is to set $V(q) = -\beta^{-1} \ln \pi(q)$, and then the marginal distribution on q from B-G will have the density function $\pi(\cdot)$.

This chapter is concerned with the following questions:

1. Although the friction parameter c does not affect the invariant distribution, it does affect the rate of convergence. How should c be chosen for faster convergence and hence accelerated sampling?

2. If sampling from B-G is the objective, the inverse temperature β does not need to be kept constant over the total simulation time T. How should the cooling schedule $t \mapsto \beta(t), t \in [0,T]$ be chosen in order to minimize the distance between the distribution of $[q(T), p(T)]$ and the desired B-G?

The short answers will be:

1. Friction should be chosen so that the local quadratic approximation to the

potential corresponds to a critical damped oscillator.

2. A near-optimal cooling schedule, which is a finite-dimensional map from integer (i.e., the step count) to real (i.e., the temperature at each step), could be obtained by minimizing an error bound, which does not (directly) depend on the dimension of the system (and hence the optimization does not become harder in high-dimensional cases), but only on a few real-valued parameters that characterize the energy landscape.

   In many cases, an inverse linear cooling schedule, which overheats the system and then inverse linearly drops the temperature to the desired value, provides a better efficiency than many other schedules.

6.2 A concise review and our contribution

Methods for sampling Boltzmann-Gibbs distribution are greatly desired due to its practical usages, partly because various phase-space integrals, which are widely used in constant temperature statistical physics [57, 108] and non-perturbative calculations of quantum field theories [163], can be approximated by summations of functions of sampled points in phase-space. This sampling is, however, a known computational challenge [24, 134, 163]. The nonlinearity of the potential and the curse of dimensionality, for instance, make sampling methods slowly convergent.

Classical sampling approaches include purely statistical methods such as Metropolis algorithm and importance sampling that are solely for sampling purposes (see for instance [25] and references therein for a review and comparison), stochastic molecular dynamics (primarily Langevin dynamics), deterministic dynamics plus an external thermostat (such as Nosé-Hoover [217, 144], Berendsen [28] or Andersen [7] thermostats), Hybrid Monte Carlo [84] (which introduces auxiliary dynamics to avoid random walks), etc. We also refer to [234] for an example that combines stochastic molecular dynamics and purely statistical approach.

Langevin dynamics, which adds friction and noise to mechanical equations to model energy exchange with a heat bath [287, 247, 248], is a good candidate for

sampling purposes. It has been shown in the context of classical molecular sampling that both stochastic dynamics and deterministic dynamics with thermostats outperform purely statistical methods in convergence rate as the size of the system grows, and stochastic dynamics has a robust behavior with respect to space dimension (we refer to [57] for a linear alkane molecule). Also, since overdamped Langevin is a special case of Hybrid Monte Carlo [63], it is not surprising to observe cases in which Langevin dynamics is computationally more efficient than purely statistical methods. Moreover, if the system is stiff or multiscale, existing stiff or multiscale Langevin integrators such as SIM [276] or FLAVOR [275] can be directly employed for accelerated computation.

Various concepts related to the idea of annealing have been proposed. A cool schedule was first introduced in Simulated Annealing algorithm [166] for global optimization, which can also be viewed as (uniformly) sampling from the set of global minimizers of V. Temperature accelerated dynamics has been proposed in [260] for events simulations. The concept there is to raise temperature of the system to make rare events occur more frequently, intercept each attempted escape from potential wells and extrapolate time to low temperature. Another temperature approach has been used to calculate free energy [189]. In that method, overheated auxiliary variables are introduced to equilibrate the collective variables faster. The type of annealing that we use is a global cooling schedule that was used in Simulated Annealing. Surprising as it may seem, we have **never** seen any demonstration of employing annealing for sampling purposes, although the abundant applications for optimization purposes are needless to mention.

In addition, although the use of Langevin for sampling is well known, to the best of our knowledge, there has been **no** study on an optimal choice of friction in Langevin dynamics.

Our proposed strategy (tuning friction and annealing temperature) is distinct from prevailing accelerated sampling methods, such as conformational flooding [125], replica exchange [269], umbrella sampling [282], self-guided MD [299], hyperdynamics [294], affine invariant ensemble sampler [122], and many others re-

viewed in [31], especially in the sense that it can be used **concurrently** with many of these methods. Indeed, while tuning friction is mostly restricted to dynamics based methods, annealing may apply to any method that involves temperature. Note that temperature is a rather general notion because it can often be introduced artificially; for instance, see [212] for an example in which temperature is introduced in an MCMC algorithm for Bayesian updating. Based on this reason, there is no point in comparing our strategy with most other acceleration approaches.

6.3 Methodology

6.3.1 Background algorithms

Although any Langevin integrator can serve as a background algorithm and be tuned and annealed, we base our numerical simulations on Geometric Langevin Algorithm (GLA) introduced in [41], which is recapped as follows:

$$\begin{cases} \hat{p}_n &= e^{-c_n h} p_n + \sqrt{\frac{1-e^{-2c_n h}}{\beta_n}} \xi_n \\ q_{n+1} &= q_n + h\hat{p}_n \\ p_{n+1} &= \hat{p}_n - h\nabla V(q_{n+1}) \end{cases} \tag{6.3}$$

where h is the timestep length, ξ_n's are i.i.d. standard normal random variables, and $c_n = c$ and $\beta_n = \beta$ in absence of friction tuning or temperature annealing.

The choice of GLA is motivated by its conformal-symplecticity and long-time properties [41]. Specifically, under certain conditions, GLA is not only pathwise accurate but also convergent towards B-G in the sense that its invariant measure deviates $O(h)$ from Boltzmann-Gibbs measure in total variation norm (and hence it captures statistical properties). It is worth mentioning that similar properties are shown to hold under weaker conditions for a Metropolized version of GLA [234, 42], which can also be tuned and annealed for accelerated samplings.

For multiscale or stiff systems (where $V(q) = V_0(q) + \epsilon^{-1}V_1(q)$ for instance), FLAVORs [275] are possible alternative background algorithms that are also conformal-symplectic (we also refer to SIMs [276] for quadratic stiff potentials).

6.3.2 Choice of friction

If V is quadratic (of the form $V = \frac{q^T K q}{2}$), we show in Section 6.4.1 that optimal acceleration is achieved by choosing $c = 2K^{\frac{1}{2}}$ so that all degrees of freedom of the harmonic oscillator are critically damped. It is also important to notice that this choice does not depend on temperature.

Based on this observation, we heuristically propose to tune the friction c_n at each time step of the simulation according to the Hessian of the potential V:

$$\begin{cases} k_n & = \begin{cases} \frac{1}{2}\frac{\partial^2 V}{\partial q^2}(q_n), & \frac{\partial^2 V}{\partial q^2}(q_n) \succ 0 \\ \alpha^2/4I, & \text{otherwise} \end{cases} \\ c_n & = 2\sqrt{k_n} \end{cases} \tag{6.4}$$

where α is a fixed real parameter, preassigned to handle the case of negative curvature; for instance, it could be equal to 0 or to the original value of c.

6.3.3 Choice of temperature

Annealing has successfully been applied to optimization problems [166]. A cooling schedule describes how to choose $T(n) = 1/\beta_n$ as a function of n. For optimization-based cooling schedules, one requires $\lim_{i\to\infty} T(i) = 0$. We refer to [127, 73, 284] for general reviews of optimization-based cooling schedules, and to [114, 132] for theoretical bounds on the convergence of specific schedules.

In this chapter, however, we are interested in situations where the total number of steps N is finite and fixed, the final temperature $T(N) = T_f = 1/\beta > 0$ is strictly positive and is the temperature at which one wishes to sample B-G.

It is then natural to seek to minimize the distance between the distribution of (q_N, p_N) and B-G at temperature $1/\beta$ using $T(1), \ldots, T(N-1)$ as optimization variables. In Section 6.4.2 we derive a bound on this distance using convergence rates of Markov chains. Notice that this bound holds in disregard of the dimension of the phase space. A numerical minimization of that bound suggests the following near-optimal cooling schedule for $T_f > 0$ (for $T_f = 0$ we refer to [73] and references

therein) and $N < N_0$ (N_0 is the number of steps needed for sampling by a naive Langevin simulation; see Section 6.4.3 for details):

$$\beta_n = \frac{n}{N}\frac{1}{T_f} + (1 - \frac{n}{N})\frac{1}{T_i}, \qquad T(n) = 1/\beta_n, \tag{6.5}$$

where N is the total-number of simulation steps, T_f the temperature at which the Gibbs distribution needs to be sampled, and the initial temperature $T_i > T_f$ is a free parameter chosen to overcome the maximal potential barrier, i.e., $T_i \gg C_V/k$ (for simplicity we let the Boltzmann constant k be equal to one in our setting; C_V can be intuitively interpreted as the maximum elevation in potential landscape, and we refer to [81] for a rigorous definition).

6.3.4 Friction and temperature accelerated sampling

Put together, annealed and tuned GLA (AnnealTuneGLA) for accelerated B-G sampling is the following:

$$\begin{cases} k_n &= \begin{cases} \frac{1}{2}\frac{\partial^2 V}{\partial q^2}(q_n) & \frac{\partial^2 V}{\partial q^2}(q_n) \succ 0 \\ \alpha^2/4 & otherwise \end{cases} \\ c_n &= 2\sqrt{k_n} \\ \beta_n &= \frac{n}{N}\frac{1}{T_f} + (1 - \frac{n}{N})\frac{1}{T_i} \\ \hat{p}_n &= e^{-c_n h} p_n + \sqrt{\frac{1-e^{-2c_n h}}{\beta_n}}\xi_n \\ q_{n+1} &= q_n + h\hat{p}_n \\ p_{n+1} &= \hat{p}_n - h\nabla V(q_{n+1}) \end{cases} \tag{6.6}$$

Compared to the background GLA, the distribution of the accelerated trajectory at a fixed time is closer to the desired B-G in the total variation sense (see numerical experiments below). A possible exact preservation of a near-by distribution is, however, not yet proved for AnnealTuneGLA. Accelerations due to tuning friction and annealing temperature are independent.

It is worth mentioning that 1st-order GLA is not unconditionally stable, nor is AnnealTuneGLA. Therefore, h or α should not be chosen to be too large.

6.4 Analysis and optimization

6.4.1 Optimal friction in linear systems

In this section, we will show that with β fixed, the choice of $c = 2\sqrt{k}$ will enable the fastest convergence of the following system:

$$\begin{cases} dq &= pdt \\ dp &= -kqdt - cpdt + \sigma dW \end{cases} \tag{6.7}$$

where $\sigma = \sqrt{2c/\beta}$. Assume k is a scalar for the moment. For our purpose, consider positive k, because if k is 0 the system decouples, and if k is negative the system is not ergodic and does not admit an invariant distribution.

The solution to the above linear system can be explicitly written as

$$\begin{cases} q(t) &= B_{11}(t)q(0) + B_{12}(t)p(0) + \int_0^t B_{12}(t-s)\sigma dW_s \\ p(t) &= B_{21}(t)q(0) + B_{22}(t)p(0) + \int_0^t B_{22}(t-s)\sigma dW_s \end{cases} \tag{6.8}$$

where $B(t)$ is the fundamental matrix defined by the autonomous ODE $\frac{dB}{dt} = \begin{bmatrix} 0 & 1 \\ -k & -c \end{bmatrix} B$, and written in block form as

$$B(t) = \begin{bmatrix} B_{11}(t) & B_{12}(t) \\ B_{21}(t) & B_{22}(t) \end{bmatrix} = \exp\left(\begin{bmatrix} 0 & 1 \\ -k & -c \end{bmatrix} t \right) \tag{6.9}$$

After calculating out the matrix exponential, the expectation of position writes as follows

$$\begin{aligned} \mathbb{E}q(t) &= B_{11}(t)q(0) + B_{12}(t)p(0) \\ &= \frac{e^{\frac{1}{2}(-c+\sqrt{c^2-4k})t}\left(c+\sqrt{c^2-4k}\right) - e^{\frac{1}{2}(-c-\sqrt{c^2-4k})t}\left(c-\sqrt{c^2-4k}\right)}{2\sqrt{c^2-4k}} q(0) \\ &+ \frac{e^{\frac{1}{2}(-c+\sqrt{c^2-4k})t} - e^{\frac{1}{2}(-c-\sqrt{c^2-4k})t}}{\sqrt{c^2-4k}} p(0) \end{aligned} \tag{6.10}$$

Naturally, the expectation approaches 0 as $t \to +\infty$. Recall that c and k are nonnegative reals. We will show in the following discussion that the maximum speed of convergence toward 0 will be achieved with $c = 2\sqrt{k}$:

1. When $c^2 - 4k > 0$, $-c - \sqrt{c^2 - 4k} < -c + \sqrt{c^2 - 4k} < 0$ and none of the coefficients are zero. Therefore the bottleneck for convergence of $B_{11}(t)$ and $B_{12}(t)$ will be $e^{\frac{1}{2}\left(-c+\sqrt{c^2-4k}\right)t}$, which will be minimized as $c^2 \downarrow 4k$.

2. When $c^2 - 4k = 0$, $B_{11} = \frac{1}{2}e^{-ct/2}(2 + ct)$ and $B_{12} = e^{-ct/2}t$.

3. When $c^2 - 4k < 0$, define a real number $\omega = \sqrt{4k - c^2}$. $B_{11} = e^{-ct/2}(c\sin(\omega t/2)/\omega + \cos(\omega t/2))$ and $B_{12} = e^{-ct/2}2\sin(\omega t/2)/\omega$. Notice $\cos(\omega t/2)$ and $\sin(\omega t/2)$ can not be simultaneously zero, and therefore the convergence rate is controlled by $e^{-ct/2}$, which will be minimized when $c^2 \uparrow 4k$.

Hence when $c = 2\sqrt{k}$ this linear system (6.7) converges the fastest. Notice that this choice corresponds to a critically damped system (as opposed to overdamped or underdamped).

Remark 6.4.1. *One may carry out the same analysis for $p(t)$, and the result will be the same (omitted). Therefore, the same choice will enable the fastest convergence towards B-G, because B-G is, after all, a distribution of q and p.*

When the system is linear but multi-dimensional, k can be assumed without loss of generality to be a symmetric matrix, and it can be immediately seen that there is no theoretical difficulty because one can diagonalize k and choose c diagonal-wisely. Therefore, any numerical method that calculates the square root of a matrix could work here for getting c. There are many possible numerical approaches for square rooting matrices, for instance by preconditioning if the matrix has some special structure (which is usually the case in molecular systems), or as in [137] or [140], but for consideration of conciseness we will not discuss this numerical topic.

6.4.2 Error bound of cooling schedules

Lemma 6.4.1 (Spectral gap and weak convergence rate)**.** *Consider an aperiodic and irreducible homogeneous Markov Chain* $(X_0, X_1, \ldots, X_n, \ldots)$*, with transition operator (matrix)* A*. Suppose* A *is diagonalizable, and the initial state* X_0 *is drawn from the distribution* μ*. Denote by* π *the invariant distribution of this Markov Chain, and by* ρ *the second largest absolute value of* A *eigenvalues, which will show to be the convergence rate. Given an arbitrary test function* f *that maps the state space* E *of the chain to* $\mathbb{R}$*, then* $\rho < 1$ *and*

$$|\mathbb{E}_\mu f(X_n) - \mathbb{E}_\pi f| \leq \rho^n |\mathbb{E}_\mu f - \mathbb{E}_\pi f| \tag{6.11}$$

Proof. For an easy illustration, write in finite dimensional linear algebra language.

Denote by $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$ the ordered eigenvalues of A. Since the chain is irreducible and aperiodic, it is well known that $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_d > -1$, and $\rho = max(a_2, |a_d|) < 1$. Since A is diagonalizable, it admits both left eigenvectors and right eigenvectors. Denote by $\psi_1, \psi_2, \ldots, \psi_d$ and $\phi_1, \phi_2, \ldots, \phi_d$ left and right eigenvectors associated to $\lambda_1, \lambda_2, \ldots, \lambda_d$. Assume without loss of generality that all ψ's have entries sum up to 1. It can be easily shown that they are biorthogonal. Eigenspace associated to eigenvalue 1 is one dimensional.

Now represent the distribution μ as a row vector a that indicates state space density. Apparently, entries of a are non-negative and sum up to 1. Moreover, it is a classical proof that ψ_1 represents the invariant distribution π. In addition, denote f by a column vector f, whose entries are arbitrary. Then (6.4.1) can be rewritten as

$$|aA^n f - \psi_1 f| \leq \rho^n |af - \psi_1 f| \tag{6.12}$$

Assume $g = f - \begin{bmatrix} \psi_1 f \\ \ldots \\ \psi_1 f \end{bmatrix}$, then since entries of a sum up to 1, the above inequality

is equivalent to

$$|aA^n g| \leq \rho^n |ag| \tag{6.13}$$

Expand a in ψ's and g in ϕ's: $a = a_1\psi_1 + \cdots + a_d\psi_d$, $g = g_1\phi_1 + \cdots + g_d\phi_d$. By the definition of g, it is easy to check that $\psi_1 g = 0$. Since $\phi_j (j \neq 1)$ is orthogonal to ψ_1, we have $g_1 = 0$. Therefore

$$\begin{aligned} |aA^n g| &= |a_1 g_1 + a_2 \lambda_2^n g_2 + \cdots + a_d \lambda_d^n g_d| \\ &= |a_2 \lambda_2^n g_2 + \cdots + a_d \lambda_d^n g_d| \\ &\leq \rho^n |a_2 + \cdots + a_d| \leq \rho^n |ag| \end{aligned} \tag{6.14}$$

□

Remark 6.4.2. *The chain does not have to be reversible (i.e., A needs not be self-adjoint). In fact, although Metropolis algorithm is reversible, Langevin dynamics generally is not.*

Remark 6.4.3. *The mild requirement that A is diagonalizable can be relieved, and a similar result will still hold, because the eigenspace associated to eigenvalue 1 is one dimensional and g_1 will always be 0. By using Jordan canonical form of A, one can prove that there exists a positive integer M so that $\|A^M\|_{Frobenius} < 1$, and therefore the ρ^n bound for the diagonalizable case could be safely replaced by $\rho^{n/N}$. Convergence might be slower, however still certain.*

Corollary 6.4.1 (Local error bound)**.** *Denote by μ_i the distribution of the phase space coordinate obtained by i steps of AnnealGLA simulation using a cooling schedule $T(\cdot)$, by $\pi_T(i)$ the Boltzmann-Gibbs distribution at temperature $T(i)$, and by ρ_i the convergence rate of the Markov Chain given by Langevin integrator at temperature $T(i)$ (TuneGLA in our case). Then in the sense of total variation,*

$$\|\mu_i - \pi_{T(i)}\|_{TV} \leq \rho_i \|\mu_{i-1} - \pi_{T(i)}\|_{TV} \tag{6.15}$$

Proof. This is a direct application of Lemma 6.4.1. Choose f to be the index function I_A for arbitrary measurable set A. Recall that total variation norm is defined as $\|\mu - \pi\|_{TV} := \sup_{A\in\mathcal{B}} |\mu(A) - \pi(A)|$, where $\mathcal{B}$ is the σ-algebra of measurable space. □

Lemma 6.4.2 (Global error bound)**.** *Let* $a_i = \|\mu_i - \pi_{T(i)}\|_{TV}$, $b_i = \|\pi_{T(i-1)} - \pi_{T(i)}\|_{TV}$, $p_i = \prod_{k=i}^{N} \rho_k$. *Then*

$$a_N \leq a_1 p_2 + \sum_{j=2}^{N} b_j p_j \tag{6.16}$$

Proof. By Corollary 6.4.1 and triangle inequality, we have

$$\|\mu_i - \pi_{T(i)}\|_{TV} \leq \rho_i \|\mu_{i-1} - \pi_{T(i)}\|_{TV} \leq \rho_i (\|\mu_{i-1} - \pi_{T(i-1)}\|_{TV} + \|\pi_{T(i-1)} - \pi_{T(i)}\|_{TV}) \tag{6.17}$$

i.e., $a_i \leq (a_{i-1} + b_i)\rho_i$. By induction, we obtain

$$a_N \leq a_1\rho_2\rho_3\ldots\rho_N + b_2\rho_2\rho_3\ldots\rho_N + b_3\rho_3\ldots\rho_N + \cdots + b_N\rho_N \tag{6.18}$$

□

Assumption 6.4.1 (Bound on spectral gap)**.** *Based on [107], [250], [80], [81], [41], [198] and transition state theory, it is reasonable to assume that in the stable regime of integration:*

$$0 < \rho_i \leq 1 - f(h, h_0)e^{-\frac{C_V}{T(i)}} \tag{6.19}$$

where h *is the integration timestep,* h_0 *is a constant indicating the stability limit on step length (and hence* $0 \leq h \leq h_0$*),* $f(a, b)$ *is some function such that* $f(0, b) = 0$ *and* $f(a, b) \leq 1$*, and* $C_V \geq 0$ *corresponds the elevation of potential, which is a constant defined as follows:*

Assumed without loss of generality that $\inf V = 0$*. Define* $V_{\min} = \{p \in E :$

$V(p) = 0\}$. *Define the elevation of a path* $\gamma(x, y)$ *linking* x *and* y *by:*

$$\text{Elev}(\gamma(x, y)) = \sup_{p \in \gamma(x,y)} V(p) \tag{6.20}$$

Define the elevation between two points by the minimal elevation path:

$$\text{Elev}(x, y) = \inf_{\gamma(x,y)} \text{Elev}(\gamma(x, y)) \tag{6.21}$$

Define the elevation of potential by:

$$C_V = \sup_{x \in E} \left(\inf_{y \in V_{\min}} (\text{Elev}(x, y) - V(x)) \right) \tag{6.22}$$

Remark 6.4.4. *This assumption is reasonable due to various reasons:*

1. *Diaconis and Stroock proved a general theorem on bounds of eigenvalues of Markov chain [81], and as a corollary we will have* $\rho \leq 1 - \frac{1}{d^3} e^{-C_V/T}$ *for a Metropolis chain at temperature* T*, where* C_V *is the one in* (6.22) *and* d *is a constant (corresponding to our* $f(h, h_0)$*). The Markov chain of AnnealTuneGLA is a more general case, and we assume the same thing still holds.*

2. *By using Freidlin-Wentzell theory [107], Schütte and Huisinga [250] show that in the low temperature limit the characteristic time of crossing a potential barrier of height* C_V *is* $\tau = e^{C_V/T}$*. Therefore by large deviation theory [80], the convergence rate* ρ *given by one step update with length of* h *will be* $\exp(-Ch/\tau)$ *for some* C*, where* C *certainly depends on* h_0*. Taylor expansion to the 1st-order turns this to our assumption when* h *is small.*

3. *The form of how the convergence rate depends on the temperature dates back to the Arrhenius rate law and the development of transition state theory [173].*

4. *GLA is geometric ergodic [41]. Our assumption implies* $0 \leq \rho < 1$*, which is consistent with this.*

Lemma 6.4.3. *$a_1 = 1$ if the initial condition of simulation is $(q_1, p_1) = x$ for an arbitrary deterministic value x.*

Proof. Notice μ_1 is a Dirac delta and $\pi_{T(1)}$ is a continuous distribution. Choose $\hat{A} = \{x\}$, then

$$\|\mu_1 - \pi_{T(1)}\|_{TV} := \sup_A |\mu_1(A) - \pi_{T(1)}(A)| \geq |\mu_1(\hat{A}) - \pi_{T(1)}(\hat{A})| = |1 - 0| = 1 \quad (6.23)$$

Since a total variation distance could at most be 1, $a_1 := \|\mu_1 - \pi_{T(1)}\|_{TV} = 1$. □

Lemma 6.4.4. *Assume $0 \leq T(j-1) - T(j) \ll T(j)$, then*

$$0 \leq b_j \leq \mathbb{E}H_{T(j)} \frac{T(j-1) - T(j)}{T(j)^2} + o\left(T(j-1) - T(j)\right) \quad (6.24)$$

where $\mathbb{E}_{T(j)}H$ is the average energy of B-G at temperature $T(j)$. Denote $\mathbb{E}_{T(j)}H = \alpha_j T(j)$, then

$$0 \leq b_j \leq \alpha_j \frac{T(j-1) - T(j)}{T(j)} + o\left(T(j-1) - T(j)\right) \quad (6.25)$$

Oftentimes $\alpha_j \approx 1$.

Proof. For conciseness, use the notation $T = T(j)$ and $dT = T(j-1) - T(j) \ll T$ throughout this proof. We have

$$b_j = \frac{1}{2} \int_{T^*Q} \left| \frac{1}{Z_T} e^{-H/T} - \frac{1}{Z_{T+dT}} e^{-H/(T+dT)} \right| dqdp \geq 0 \quad (6.26)$$

where

$$\begin{aligned} Z_{T+dT} &= \int_{T^*Q} e^{-H/(T+dT)}\, dqdp \\ &= \int_{T^*Q} e^{-H/T} e^{-H\left(-dT/T^2 + o(dT)\right)}\, dqdp \\ &= \int_{T^*Q} e^{-H/T} \left(1 + HdT/T^2 + o(dT)\right)\, dqdp \\ &= Z_T \left(1 + \mathbb{E}_T H dT/T^2 + o(dT)\right) \end{aligned} \quad (6.27)$$

Therefore,

$$\begin{aligned}
b_j &= \frac{1}{2}\int_{T^*Q}\frac{1}{Z_T}e^{-H/T}\left|1-\frac{Z_T}{Z_{T+dT}}e^{-H(1/(T+dT)-1/T)}\right|\,dqdp \\
&= \frac{1}{2}\int_{T^*Q}\frac{1}{Z_T}e^{-H/T}\left|1-\frac{1}{1+\mathbb{E}_T HdT/T^2+o(dT)}e^{-H\left(-dT/T^2+o(dT)\right)}\right|\,dqdp \\
&= \frac{1}{2}\int_{T^*Q}\frac{1}{Z_T}e^{-H/T}\left|1-\left(1-\mathbb{E}_T HdT/T^2+o(dT)\right)\left(1+HdT/T^2+o(dT)\right)\right|\,dqdp \\
&= \frac{1}{2}\int_{T^*Q}\frac{1}{Z_T}e^{-H/T}\left|\mathbb{E}_T HdT/T^2-HdT/T^2+o(dT)\right|\,dqdp \\
&\leq \frac{1}{2}\int_{T^*Q}\frac{1}{Z_T}e^{-H/T}\left(\left|\mathbb{E}_T HdT/T^2\right|+\left|HdT/T^2\right|+o(dT)\right)\,dqdp \qquad (6.28)
\end{aligned}$$

Since H is bounded from below, assume without loss of generality that $H \geq 0$. Then

$$\begin{aligned}
b_j &\leq \frac{1}{2}\int_{T^*Q}\frac{1}{Z_T}e^{-H/T}\left(\mathbb{E}_T HdT/T^2+HdT/T^2+o(dT)\right)\,dqdp \\
&= \mathbb{E}_T HdT/T^2+o(dT) \qquad (6.29)
\end{aligned}$$

□

Remark 6.4.5. *This bound is tight in the following sense:*

Lemma 6.4.5. *If H is a positive definite quadratic function (for instance, normalized harmonic oscillator $H(q,p)=(q^2+p^2)/2$), then $\alpha_j = 1$, and the bound of b_j in Lemma 6.4.4 is reached: $b_j = \alpha_j \frac{T(j-1)-T(j)}{T(j)} + o\left(T(j-1)-T(j)\right)$.*

Proof. Notice that Boltzmann-Gibbs in this case (possibly after linear transformation in phase space) can be written as $d\mu = e^{-H(q,p)/T}/TdH$. $\alpha_j = 1$ hence obviously follows.

For the latter, set $H_c = \log\frac{T_1}{T_2}\frac{T_1T_2}{T_1-T_2}$ by solving $\frac{1}{T(j-1)}e^{-H_c/T(j-1)}-\frac{1}{T(j)}e^{-H_c/T(j)} = 0$ so that sign changes across H_c. Since $T(j-1)\geq T(j)$ and $\int_{H_c}^{\infty} = 1-\int_0^{H_c}$, we

have

$$\begin{aligned} b_j &= \int_0^{H_c} \left(\frac{1}{T(j-1)} e^{-H/T(j-1)} - \frac{1}{T(j)} e^{-H/T(j)} \right) dH \\ &= \frac{T(j-1)}{T(j)}^{-\frac{T(j)}{T(j-1)-T(j)}} - \frac{T(j-1)}{T(j)}^{-\frac{T(j-1)}{T(j-1)-T(j)}} \end{aligned} \tag{6.30}$$

As $T(j-1) - T(j) \to 0$, L'Hospital's rule gives

$$b_j \to \frac{T(j-1) - T(j)}{T(j)} \tag{6.31}$$

□

Theorem 6.4.1. *Assume a Langevin integrator $(q_{i-1}, p_{i-1}) \mapsto (q_i, p_i)$ at temperature $T(i)$ with step length h has a bound on convergence rate $\rho_i \leq 1 - f(h, h_0) e^{-\frac{C_V}{T(i)}}$ (Assumption 6.4.1), corresponds to a Markov Chain with a diagonalizable transition operator, the initial condition is $(q_1, p_1) = x$ for some deterministic x, and the cooling schedule satisfies $0 \leq T(j-1) - T(j) \ll T(j)$, then*

$$\begin{aligned} \|\mu_N - \pi_N\|_{TV} \leq \sum_{j=2}^{N} \left(\alpha_j \frac{T(j-1) - T(j)}{T(j)} \prod_{k=j}^{N} \left(1 - f(h, h_0) e^{-\frac{C_V}{T(k)}} \right) + o\left(T(j-1) - T(j)\right) \right) \\ + \prod_{k=2}^{N} \left(1 - f(h, h_0) e^{-\frac{C_V}{T(k)}} \right) \end{aligned} \tag{6.32}$$

where μ_N is the distribution of (q_N, p_N), π_N is Boltzmann-Gibbs at temperature $T(N)$, and $\alpha_j = \mathbb{E}_{T(j)} H / T(j)$ (oftentimes $\alpha_j \approx 1$).

Proof. By applying Lemma 6.4.3, Lemma 6.4.4, Assumption 6.4.1 to Lemma 6.4.2.

□

6.4.3 Optimization with respect to cooling schedules

Naturally, one would like to minimize the error bound (6.32) with respect to $T(n)$'s. This is however difficult because (6.32) is a highly nonlinear function of T. Instead, we consider the following prevailing types of cooling schedules (denote by T_f the

final temperature at which we want to sample the B-G, and by N the number of steps we can afford to employ):

Inverse logarithmic cooling:

$$T(n) = T_f \frac{\log(N+1)}{\log(n+1)} \tag{6.33}$$

This is the most popular schedule for optimization ([114, 132], for instance, have been frequently cited), but truncated at T_f before $T \to 0$. Recall inverse logarithmic cooling is $T(n) = \frac{C}{\log(n+1)}$, and C is fixed by requiring $T(N) = T_f$. When N is fixed, there is no need to choose any parameter. This schedule will serve as our benchmark.

Shifted inverse logarithmic cooling:

$$T(n) = T_f + \frac{C}{\log(n+1)} \tag{6.34}$$

where $C > 0$ is the free parameter to be optimized. $T(N)$ is set to be T_f.

Exponential cooling:

$$T(n) = T_f e^{\tilde{C}(N-n)} = T_f C^{N-n} \tag{6.35}$$

where $C = e^{\tilde{C}} > 1$ is the free parameter to be optimized.

Shifted exponential cooling:

$$T(n) = T_f + \tilde{C} \cdot C^{-n} \tag{6.36}$$

where $\tilde{C} > 0$ and $C > 1$ are free parameters. For ease on optimization, we chose $\tilde{C} = 10^{-4} T_f C^N$ so that temperatures 'smoothly' cool to T_f, and are left to optimize only one free parameter.

Linear cooling:

$$T(n) = \frac{n}{N}T_f + (1 - \frac{n}{N})T_i \tag{6.37}$$

where $T_i > T_f$ is the free parameter. This is used in [270] for optimization purposes. This seemingly too fast cooling schedule does give a small error bound in typical cases (see below).

Inverse linear cooling:

$$T(n) = 1/\left(\frac{n}{N}\frac{1}{T_f} + (1 - \frac{n}{N})\frac{1}{T_i}\right) \tag{6.38}$$

where $T_i > T_f$ is the free parameter. Instead of linearly interpolating the temperature, this linearly interpolates β which is the inverse of temperature to ensure more steps at low temperatures.

Optimal error bound: We optimize the error bound (6.32) for different total numbers of steps (N's) with respect to the cooling schedules (or more precisely, their parameters) described above. As indicated in Table 6.1, the optimal choice depends on the total simulation time (N, or more precisely, the ratio between Nh and the mixing time of the original system). Unless N is too small or too large, optimal inverse linear cooling produces a small error bound, optimal linear and exponential coolings have close performances as well, and all three optimal cooling schedules are similar. If the number of steps is too small, B-G will not be approximated well by any cooling schedule, and it is better to use the trivial schedule of constant temperature. If the number is instead too large (usually not the case of interest because accelerated sampling is desired), most types of cooling schedules will yield small errors, and surprisingly, shifted exponential cooling outperforms inverse logarithmic cooling, which is a popular cooling schedule for large N.

In these experiments, $T_f = 20$, $C_V = 150$, $f(h, h_0) = 1$, and $\alpha_j = 1$. In this typical setting T_f/C_V is small and the B-G distribution is concentrated in potential wells, h is close to h_0, and $\alpha_j \approx 1$. If the T_f/C_V is large, however, the optimization

N	Constant (no cooling)	Inverse log (benchmark)	Shifted inverse log	Exp	Shifted exp	Linear	Inverse linear
200	0.896	1.304	0.950	**0.896**[1]	**0.896**[1]	**0.896**[1]	**0.896**[1]
600	0.718	0.560	0.752	0.372[2]	0.718	**0.365**[2]	0.368[2]
1000	0.575	0.325	0.597	0.266[3]	0.346	0.267[3]	**0.265**[3]
2000	0.331	**0.142**	0.336	0.153[4]	0.161	0.155[4]	0.151[4]
5000	0.063	0.047	0.064	0.046[5]	**0.028**	0.047[5]	0.046[5]

[1]: Achieved by the limiting case of almost constant temperature

[2,3,4,5]: Achieved by almost the same linear-alike optimizers within each row

Table 6.1: Optimal error bound for different cooling schedules given N total steps. Within each row, bold indicates the minimum error bound. Different values of N are chosen to represent regimes of very small, small, medium, large, very large N values, in the sense of being compared to the total mixing steps which in this case renders the error bound 0.5 with a constant cooling and is $N \approx 1250$.

suggests not to anneal (result not shown). Optimization is done using MATLAB command 'fmincon'.

It is worth mentioning again that these optimization results do not directly depend on the dimension of the system nor details of the potential landscape, but only rely on several characteristic parameters: C_V, $f(h, h_0)$, and α_j.

6.5 Numerical experiments

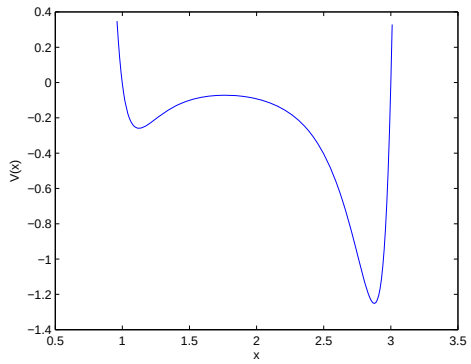


Figure 6.1: Potential energy landscape.

Consider a one dimensional nonlinear molecular system consisting of two distinct heavy (fixed) atoms and a light atom between them. It is modeled as a single degree of freedom Hamiltonian system with a Lennard-Jones potential function $V(q) = \left(q^{-12} - q^{-6}\right) + 5\left((4-q)^{-12} - (4-q)^{-6}\right)$ (Figure 6.1). The energy landscape consists of a local potential barrier and two potential wells. The attraction due to the right atom is larger than the left one. If one starts the dynamics with zero initial momentum and position in the left basin, the asymptotic (long time) position distribution will be a marginal of B-G and concentrated in the right basin. Therefore, the expectation of position q at a fixed time can be used as an indicator of the convergence rate for this nonlinear system (see also Remark 6.4.1 for why).

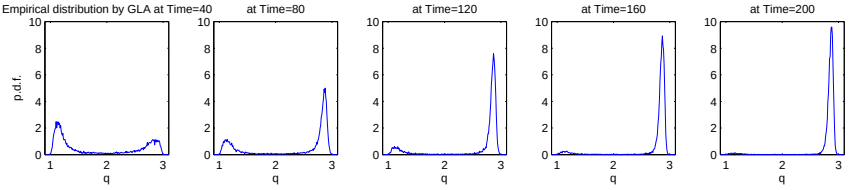


Figure 6.2: Evolution of the empirical distribution obtained by GLA (Eq. 6.3) with $c = 0.1$. The Markov process is converging as the distribution peaks more and more in the right potential basin. Simulation is done with a step length $h = 0.01$ and distributions are approximated empirically by an ensemble of 10000 trajectories.

Throughout this section we use parameters $\beta = 10$, $q(0) = 1.1$ and $p(0) = 0$. With an arbitrarily chosen $c = 0.1$, Langevin dynamics integrated with a B-G preserving method GLA (Eq. 6.3) takes more than 200 time units before indiscernible convergence (Figure 6.2).

6.5.1 Effect of friction

Enumerating c values for fixed β (and hence temperature T), one obtains different values of $\mathbb{E}[q(\text{TotalTime})]$ for a fixed total simulation time (Figure 6.3). This confirms that the value of c affects the convergence rate. The optimal fixed value is $c = 0.7$ in this example.

Although in practice it is rarely the case that an optimization can be carried out beforehand to determine the best value of c for fastest convergence of GLA, we

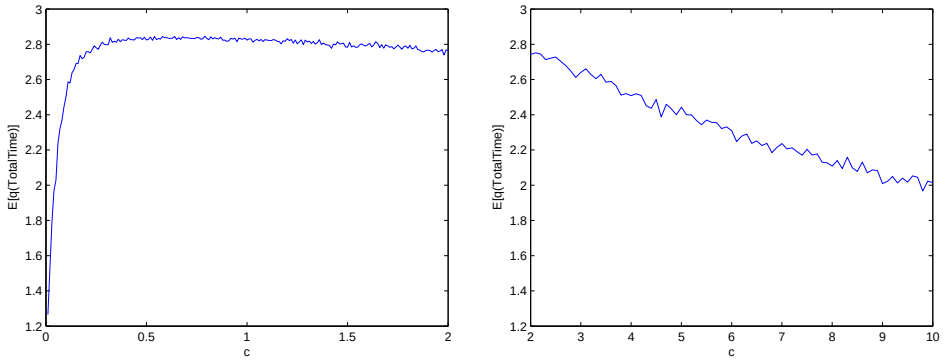


Figure 6.3: Expectations of position at a fixed time for different frictions obtained by GLA (Eq. 6.3). Larger expectation implies better convergence in this problem, and therefore this indicates the relationship between choice of c and convergence rate. The fixed time is TotalTime=100, step length is $h = 0.01$, expectations are calculated by an empirical average over an ensemble of 1000 trajectories. c values are enumerated from 0.01, 0.02, ..., 1.99, 2.00 and 2.10, ..., 19.90, 20.00.

nevertheless use GLA with the optimal friction $c = 0.7$ for comparison purposes. We will show that TuneGLA outperforms even this optimized GLA, demonstrating that c really needs to be tuned locally.

6.5.2 Additive effects of tuning friction and annealing temperature

In Figure 6.4, GLA with $c = 0.7$ (the optimal fixed value), TuneGLA (GLA with friction tuning) which adaptively tunes c but does not anneal (Eq. 6.6), AnnealGLA (GLA with temperature annealing) which uses an inverse linear cooling schedule ($C = 10T_f$) but does not tune c (Eq. 6.38), and AnnealTuneGLAs that tune and anneal with $\alpha = 0$ and $\alpha = 0.7$, respectively, are compared. We observe that tuning friction and annealing temperature individually accelerates the convergence, and their effects are additive. Therefore, the proposed AnnealTuneGLA has the fastest rate of convergence. In addition, here the choice of $\alpha = 0$ slightly outperforms $\alpha = 0.7$, which is set to be the value of the optimal c. The optimal choice of α has not been investigated.

6.5.3 Numerical validation on choices of cooling schedule

These cooling schedules have been implemented on the concrete example in Section 6.5. We did not optimize cooling schedules with respect to free parameters but used a heuristic/generic constant instead. Error on the empirical expectation of position has been investigated for each schedule in Figure 6.5. The ranking of different types of schedules depends on total simulation time and agrees with theoretical prediction (except for large total simulation times which are dominated by numerical error accumulation).

In addition to Figure 6.5 and the above discussion that compare cooling schedules for different total simulation times, we fix total time and show time dependent errors of different schedules in Figure 6.6. Here total simulation time is 30 and we are in the medium N regime. Inverse linear cooling indeed has better performances, followed closely by linear cooling, both consistent with the theoretical analysis. Rigorously speaking, one should compare cooling schedules only towards the end of the simulation, because different cooling schedules are at different temperatures in the middle of the simulation; however, the superiority of inverse linear cooling is in fact exhibited throughout the simulation.

These numerical experiments and theoretical bounds indicate that inverse linear cooling is ranked at the top. It is worth pointing out that although annealing accelerates convergence significantly, one has to choose a priori parameters (in most of our cases, total simulation step N and constant C or T_i). This issue usually needs a case-by-case investigation, but C_V (if known) could be used in conjunction with the error bound to determine N and C.

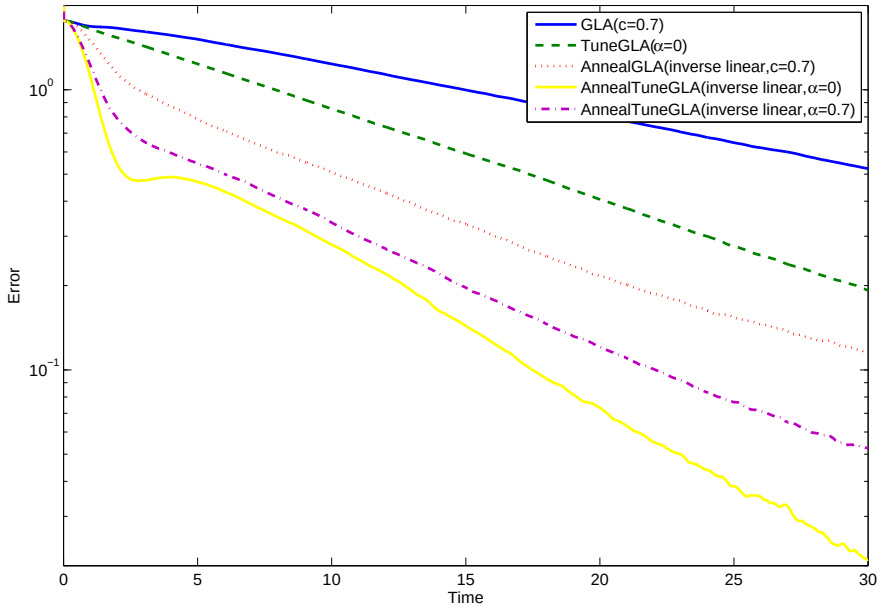


Figure 6.4: Comparison of errors of GLA, TuneGLA with c adaptively tuned, AnnealGLA with inverse linear cooling schedule, and AnnealTuneGLA with both. $c = 0.7$ that ensures fastest GLA convergence (Figure 6.3) is used in GLA and AnnealGLA. A comparison between choices of α (which indicates the value of c when curvature of potential is negative for tuning (Eq. 6.6) is also presented. Total simulation time=30 is fixed, and error at each step throughout the simulation is recorded. Simulation step length is $h = 0.01$. Error at time t is calculated by $|\frac{1}{M}\sum_{i=1}^{M} q^i(t) - \mathbb{E}q(\infty)|$, where $M = 10000$ is the total number of independent trajectories, $q^i(t)$ is the position of the ith trajectory at time t, and $\mathbb{E}q(\infty)$ is well approximated by empirical average of an ensemble of 20000 GLA trajectories at total simulation time of 300. The constant of initial temperature in the inverse linear cooling (Eq. 6.38) is $C = 10T_f$ and applies to all three AnnealGLAs.

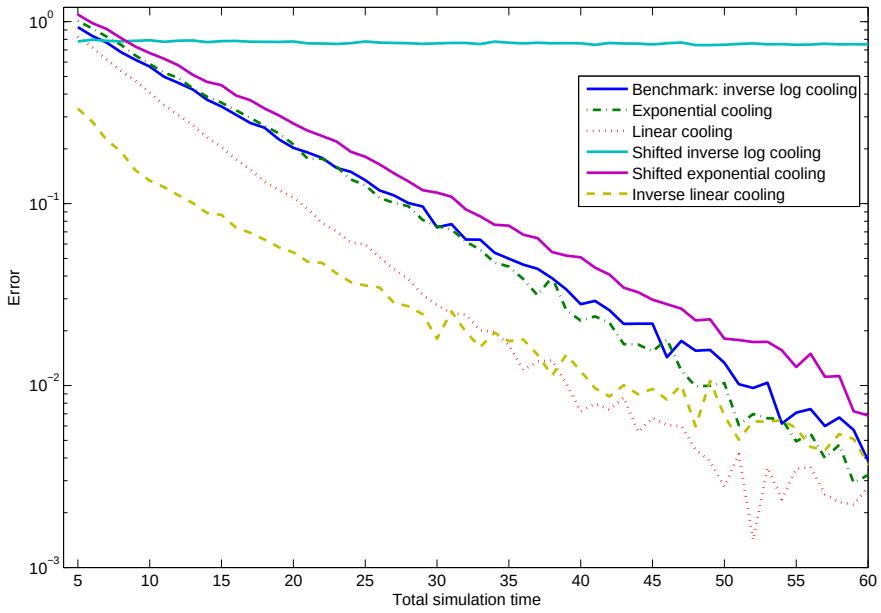


Figure 6.5: Errors of representative cooling schedules as functions of total simulation time (hence of total simulation step N too). Errors are calculated by $|\frac{1}{M}\sum_{i=1}^{M} q_N^i - \mathbb{E}q(\infty)|$, where $M = 10000$ is the total number of independent trajectories, q_N^i is the Nth step position of the ith trajectory, $N \cdot h$ is the total simulation time and the step length $h = 0.01$. $\mathbb{E}q(\infty)$ is well approximated by empirical average of an ensemble of 20000 TuneGLA trajectories at total simulation time of 300. Constants used in cooling schedules are: Shifted inverse log: $C = 0.01T_f$, Exp: $C = 1.5$, Shifted exp: $T(1) = 2T_f$, Linear: $C = 2T_f$, Inverse linear: $C = 10T_f$. Basically all settings are the same as in Section 6.5 except for total simulation time and cooling schedule used. Total simulation time is enumerated from 5 to 100 with an increment of 1.

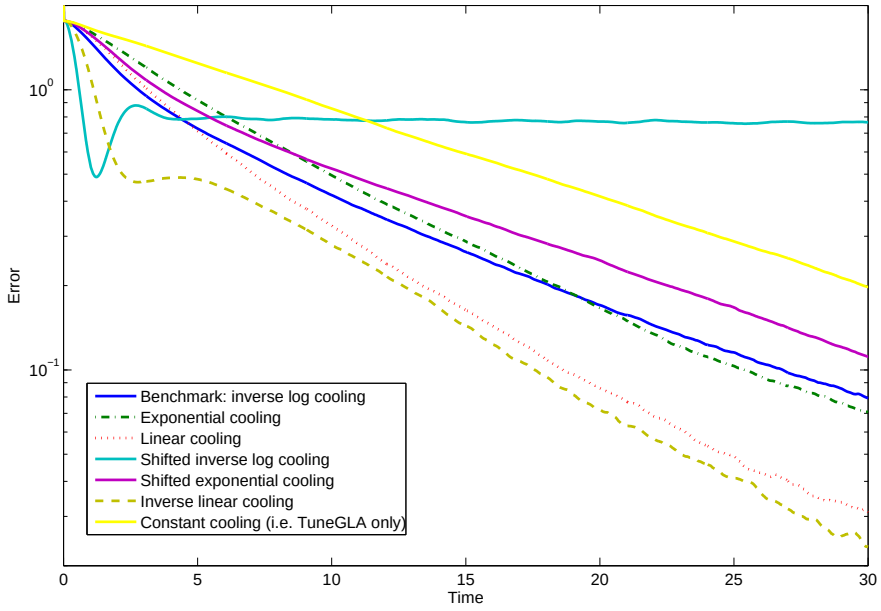


Figure 6.6: Comparison of errors of TuneGLA with c adaptively tuned and AnnealTuneGLA with different cooling schedules. Again, TuneGLA uses $\alpha = 0.7$, total simulation time=30 is fixed, and all other settings are the same as in Figure 6.5 and 6.4 too.

Chapter 7

Applications and related projects

7.1 Variational integrators for noisy multiscale circuits

When simulating the dynamics of an electrical circuit, one is faced with four difficulties: (i) the system involves external (control) forcing through external (controlled) voltage sources and resistors; (ii) the system is constrained via the Kirchhoff current (KCL) and voltage laws (KVL); (iii) the corresponding Lagrangian is degenerate; (iv) circuits in reality are always noisy and exhibiting multiple timescales. Variational integrators that collaborators and I proposed not only could overcome the first three difficulties, but also have nice structure preservation properties, including a better energy behavior and a preservation of frequency spectrum. Moreover, they could be extended to simulate noisy circuits (via the approach of stochastic variational integrator [40]) and multiscale circuits (via the approach of FLAVORs (Chapter 2)).

Most results in this section are excerpts or paraphrases of the content of a submitted manuscript [218]. In order to credit collaborators' contributions, many concepts described there will not be included in my thesis, such as the modeling and the underlying geometry (the Dirac structure). A mathematical formulation of the problem, however, is necessary and will be recapped.

7.1.1 Constrained variational formulation

A circuit, when modeled in a geometric context (we refer to [218] for details), corresponds to a mechanical system with Lagrangian $\mathcal{L} : TQ \to \mathbb{R}$ defined as

$$\mathcal{L}(q, v) = \frac{1}{2} v^T L v - \frac{1}{2} q^T C q \tag{7.1}$$

with $q(t)$ and $v(t)$ being time-dependent charges and the currents (vector) of the circuit elements. $q(t) \in Q$, where our configuration space Q will be called the charge space.

$L = \text{diag}(L_1, \ldots, L_n)$ and $C = \text{diag}\left(\frac{1}{C_1}, \ldots, \frac{1}{C_n}\right)$ are matrices that correspond to all inductors and capacitors. In the case where no inductor (resp. no capacitor) is on branch i, the corresponding entry L_i (resp. $\frac{1}{C_i}$) in the matrix L (resp. C) is zero. In the presence of mutual inductors rather than self inductors, the matrix L is not diagonal anymore, but always positive semi-definite. If not explicitly mentioned the following theory and construction is also valid for mutual inductors.

The Legendre transform $\mathbb{F}\mathcal{L} : TQ \to T^*Q$ is defined by

$$\mathbb{F}\mathcal{L}(q, v) = (q, \partial\mathcal{L}/\partial v) = (q, Lv). \tag{7.2}$$

Note that the Lagrangian can be degenerate if the Legendre transform is not invertible, i.e., L is singular.

The Lagrangian force of the system consists of a damping force that results from the resistors and an external force being the voltage sources

$$f_L(q, v, t) = -\text{diag}(R)v + \text{diag}(\mathcal{E})u \tag{7.3}$$

with $R = (R_1, \ldots, R_n)^T$ and $\mathcal{E} = (\epsilon_1, \ldots, \epsilon_n)^T$ respectively corresponding to resistors and voltage sources. If no resistor is on branch i, the corresponding entry R_i in the vector R is zero. Similarly, for the entries of the vector $\mathcal{E}$, it holds $\epsilon_i = 0$ if no voltage source is on branch i, and here we assume that the time evolution of the voltage sources is given as a time-dependent function $u_s(t)$.

The *constraint flux linkage subspace*[1] is defined by the Legendre transformation as

$$P = \mathbb{F}\mathcal{L}(\Delta_Q) \subset T^*Q,$$

where $\Delta_Q \subset TQ$ (i.e., $\Delta_Q(q) \subset T_qQ$ for all q) is a distribution that forms the Kirchhoff Current Law (KCL) constraint submanifold, which coincides with $ker(K^T)$ for a Kirchhoff Constraint matrix $K \in \mathrm{M}_{\mathbb{R}}(n, m)$ that represents the topology of the circuit.

Δ_Q^0 (the annihilator of Δ_Q) can be expressed by the image of K. Choosing another matrix $K_2 \in \mathrm{M}_{\mathbb{R}}(n, n-m)$ such that $ker(K_2^T) = im(K)$, this annihilator describes the Kirchhoff Voltage Law (KVL) constraint submanifold by

$$\Delta_Q^0(q) = \{u \in T_q^*Q \,|\, K_2^T u = 0\} \subset T_q^*Q$$

To derive the equations of motion for the circuit system, we make use of the Lagrange-d'Alembert-Pontryagin principle, i.e., search for curves $q(t)$, $v(t)$ and $p(t)$ fulfilling

$$\delta \int_0^T \mathcal{L}(q(t), v(t)) + \langle p(t), \dot{q}(t) - v(t) \rangle \, dt + \int_0^T f_L(q(t), v(t), t) \cdot \delta q(t) \, dt = 0 \quad (7.4)$$

with fixed initial and final variations $\delta q(0) = \delta q(T) = 0$ and constrained variations $\delta q \in \Delta_Q(q)$.

Taking variations gives us

$$\int_0^T \left[\left\langle \frac{\partial \mathcal{L}}{\partial q} + f_L, \delta q \right\rangle - \langle \dot{p}, \delta q \rangle + \langle \delta p, \dot{q} - v \rangle + \left\langle \left(\frac{\partial \mathcal{L}}{\partial v} \right) - p, \delta v \right\rangle \right] dt = 0 \quad (7.5)$$

for arbitrary variations δv and δp, $K^T v = 0$ and constrained variations $\delta q \in \Delta_Q(q)$.

[1] also denoted by the set of primary constraints

This leads to the constrained Euler-Lagrange equations

$$\frac{\partial \mathcal{L}}{\partial q} - \dot{p} + f_L \in \Delta_Q^0(q) \tag{7.6a}$$

$$\dot{q} = v \tag{7.6b}$$

$$\frac{\partial \mathcal{L}}{\partial v} - p = 0 \tag{7.6c}$$

$$K^T v = 0. \tag{7.6d}$$

For the Lagrangian (7.1) and the forces (7.3), the constrained Euler-Lagrange equations are

$$\dot{p} = -Cq - \operatorname{diag}(R)v + u_s + K\lambda \tag{7.7a}$$

$$\dot{q} = v \tag{7.7b}$$

$$p = Lv \tag{7.7c}$$

$$K^T v = 0, \tag{7.7d}$$

where $\lambda \in \mathbb{R}^m$ is a Lagrange multiplier.

7.1.2 Reduced variational formulation

Now we perform a reduction to project onto the constrained manifold and get rid of the Lagrange multiplier. Specifically, instead of treating the KCL as extra constraint in the form $K^T v = 0$, we directly involve the KCL form $K_2\tilde{v} = v$ with $\tilde{v} \in T_q M \subseteq \mathbb{R}^{n-m}$ for the definition of the new Lagrangian system.

Since K is constant, the constraints are integrable, i.e., the configurations q are constrained to be in the submanifold

$$C = \{q \in Q \,|\, K^T q = 0\}$$

for consistent initial values $q_0 \in C$.

For a subclass of circuits, the degeneracy of the Lagrangian will be canceled in the reduced Lagrangian. Moreover, the reduction is geometrically intrinsic

in the sense that there will be a reduced Lagrangian defined on a mesh space $TM \subseteq \mathbb{R}^{2(n-m)}$. More precisely, it holds that $T_q C = \Delta_Q(q)$ and the branch charges q can be expressed by the mesh charges $\tilde{q} \in M \subseteq \mathbb{R}^{n-m}$ as $q = K_2\tilde{q}$.

We define the constrained Lagrangian $\mathcal{L}^M : TM \to \mathbb{R}$ via pullback as $\mathcal{L}^M := K_2^*\mathcal{L} : TM \to \mathbb{R}$ with

$$\mathcal{L}^M(\tilde{q}, \tilde{v}) = \mathcal{L}(K_2\tilde{q}, K_2\tilde{v}) = \frac{1}{2}\tilde{v}^T K_2^T L K_2 \tilde{v} - \frac{1}{2}\tilde{q}^T K_2^T C K_2 \tilde{q} \tag{7.8}$$

with the Legendre transformation $\mathbb{F}\mathcal{L}^M : TM \to T^*M$ being

$$\mathbb{F}\mathcal{L}^M(\tilde{q}, \tilde{v}) = (\tilde{q}, \partial\mathcal{L}^M/\partial\tilde{v}) = (\tilde{q}, K_2^T L K_2 \tilde{v}).$$

Depending on the inductor matrix L and the circuit topology, the matrix $K_2^T L K_2$ can still be singular, i.e., the Lagrangian system can still be degenerate. We refer to Proposition 1 in [218] for more details.

The cotangent bundle T^*M is given by

$$\begin{aligned} T^*M &= \{(\tilde{q}, \tilde{p}) \in \mathbb{R}^{n-m,n-m} \,|\, (\tilde{q}, \tilde{p}) = \mathbb{F}\mathcal{L}^M(\tilde{q}, \tilde{v}) \,\text{with}\, (\tilde{q}, \tilde{v}) \in TM\} \\ &= \{(\tilde{q}, \tilde{p}) \in \mathbb{R}^{n-m,n-m} \,|\, (\tilde{q}, \tilde{p}) = (\tilde{q}, K_2^T p) \,\text{with}\, p \in P\}. \end{aligned}$$

Thus, the constrained force f_L^M in T^*M is defined as

$$f_L^M(\tilde{q}, \tilde{v}, t) = K_2^T f_L(K_2\tilde{q}, K_2\tilde{v}, t) = -K_2^T \mathrm{diag}(R) K_2 \tilde{v} + K_2^T u_s(t). \tag{7.9}$$

With $\tilde{p} \in T_{\tilde{q}}^* M \subset \mathbb{R}^{n-m}$ given as $\tilde{p} = K_2^T p$ we obtain the following reduced Lagrange-d'Alembert-Pontryagin principle:

$$\delta \int_0^T \mathcal{L}^M(\tilde{q}(t), \tilde{v}(t)) + \left\langle \tilde{p}(t), \dot{\tilde{q}}(t) - \tilde{v}(t) \right\rangle \, dt + \int_0^T f_L^M(\tilde{q}(t), \tilde{v}(t), t) \cdot \delta\tilde{q}(t) \, dt = 0 \tag{7.10}$$

with fixed initial and final variations $\delta\tilde{q}(0) = \delta\tilde{q}(T) = 0$. Taking variations gives

us

$$\int_0^T \left[\left\langle \frac{\partial \mathcal{L}^M}{\partial \tilde{q}} + f_L^M, \delta\tilde{q} \right\rangle - \langle \dot{\tilde{p}}, \delta\tilde{q} \rangle + \langle \delta\tilde{p}, \dot{\tilde{q}} - \tilde{v} \rangle + \left\langle \left(\frac{\partial \mathcal{L}^M}{\partial \tilde{v}} \right) - \tilde{p}, \delta\tilde{v} \right\rangle \right] dt = 0 \tag{7.11}$$

for arbitrary variations $\delta\tilde{v}$ and $\delta\tilde{p}$ and $\delta\tilde{q}$. This results in the reduced Euler-Lagrange equations

$$\frac{\partial \mathcal{L}^M}{\partial \tilde{q}} - \dot{\tilde{p}} + f_L^M = 0 \tag{7.12a}$$

$$\dot{\tilde{q}} = \tilde{v} \tag{7.12b}$$

$$\frac{\partial \mathcal{L}^M}{\partial \tilde{v}} - \tilde{p} = 0. \tag{7.12c}$$

For the Lagrangian (7.8) and the forces (7.9), the constrained Euler-Lagrange equations are

$$\dot{\tilde{p}} = K_2^T \left(-CK_2\tilde{q} - \mathrm{diag}(R)K_2\tilde{v} + u_s \right) \tag{7.13a}$$

$$\dot{\tilde{q}} = \tilde{v} \tag{7.13b}$$

$$\tilde{p} = K_2^T L K_2 \tilde{v}. \tag{7.13c}$$

where KVL and KCL are intrinsically satisfied. System (7.13) is a DAE system with differential variables $\tilde{q}$ and $\tilde{p}$ and algebraic variables $\tilde{v}$. The algebraic equation (7.13c) is the Legendre transformation of the system. If this is invertible (i.e., the matrix $K_2^T L K_2$ is regular), the algebraic variable v can be eliminated. In this case, the Euler-Lagrange equations (7.13) represent a non-degenerate Lagrangian system.

The equivalency between the original system (7.6) and the reduced system (7.12) is shown by Theorem 1 in [218].

7.1.3 Discrete variational principles

Due to the equivalency between the original system (7.6) and the reduced system (7.12), we can directly discretize the reduced variational principle to simulate the full circuit. For instance, replace the reduced Lagrange-d'Alembert-Pontryagin principle (7.10) by a discrete version

$$\delta\left\{h\sum_{k=0}^{N-1}\left(\mathcal{L}^M(\tilde{q}_k,\tilde{v}_k)+\left\langle\tilde{p}_k,\frac{\tilde{q}_{k+1}-\tilde{q}_k}{h}-\tilde{v}_k\right\rangle\right)\right\}+h\sum_{k=0}^{N-1}f_L^M(\tilde{q}_k,\tilde{v}_k,t_k)\delta\tilde{q}_k=0, \tag{7.14}$$

After plotting in the Lagrangian defined in (7.8) and the Lagrangian forces defined in (7.9), we obtain the updating rule (for each k-th step):

$$\begin{cases}\dfrac{\tilde{p}_k-\tilde{p}_{k-1}}{h} & =K_2^T\left(-CK_2\tilde{q}_k-\mathrm{diag}(R)K_2\tilde{v}_k+u_s(t_k)\right)\\ \dfrac{\tilde{q}_k-\tilde{q}_{k-1}}{h} & =\tilde{v}_{k-1}\\ K_2^TLK_2\tilde{v}_k & =\tilde{p}_k\end{cases} \tag{7.15}$$

with

$$\tilde{p}_0=K_2^TLK_2\tilde{v}_0 \tag{7.16}$$

It can be easily shown that this implicit updating rule has a unique solution if $K_2^T(L+h\mathrm{diag}(R))K_2$ is regular. In other words, the intrinsic degeneracy of $K_2^TLK_2$ being singular could be bypassed by the numerical integrator in many cases.

More discretization schemes and corresponding integrators can be found in [218].

7.1.4 Preservation of frequency spectrum and other structures

A peculiar observation that was not mentioned about symplectic integration elsewhere (to our knowledge) is that the frequency spectrum of the discrete solutions is much better numerically preserved by variational integrators than by non-symplectic integrators (see experiments in Section 7 in [218]). We provide a first

step in an analytical demonstration of this phenomenon:

Consider a one-degree-of-freedom oscillatory linear system (which can be shown must be a harmonic oscillator). We demonstrate that a symplectic method preserves the frequency spectrum of this system by two steps: (i) We show that for a convergent scheme the update matrix A has two eigenvalues both of norm 1 if and only if the update scheme is symplectic. (ii) We show that methods defined by matrices with norm 1 eigenvalues preserve the frequency spectrum defined on different time spans.

(i) "$\Leftarrow$": Assume the scheme defined by A is symplectic, then $\det(A) = 1$ (see, e.g., [194]). It follows with λ_1 complex conjugate to λ_2 ($\lambda_2 = \lambda_1^*$): $1 = \det(Q) \cdot \det(V) \cdot \det(Q^{-1}) = \lambda_1 \cdot \lambda_2 = |\lambda_1|^2 = |\lambda_2|^2$ and thus $|\lambda_i| = 1$, $i = 1, 2$. "$\Rightarrow$": Assume A has two complex conjugate eigenvalues $\lambda_1 = \lambda_2^*$ with $|\lambda_1| = |\lambda_2| = 1$, i.e., we write $\lambda_1 = e^{i\theta}$ and $\lambda_2 = e^{-i\theta}$ with $\theta \in \mathbb{R}$ and $V = \mathrm{diag}(e^{i\theta}, e^{-i\theta})$. Note that θ depends on the constant time step h that is used for the discretization. Let $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ be the canonical symplectic form and introduce the non-canonical symplectic form $\tilde{J} = Q^T J Q$. We show that V preserves $\tilde{J}$, and therefore A preserves J, i.e., A is symplectic. Since J is skew-symmetric with zero diagonal, $\tilde{J}$ is of the form $\begin{pmatrix} 0 & \triangle \\ -\triangle & 0 \end{pmatrix}$ with $\triangle \in \mathbb{R}$. It follows

$$
\begin{aligned}
V^T \tilde{J} V &= \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \begin{pmatrix} 0 & \triangle \\ -\triangle & 0 \end{pmatrix} \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \\
&= \begin{pmatrix} 0 & e^{i\theta} e^{-i\theta} \triangle \\ -e^{-i\theta} e^{i\theta} \triangle & 0 \end{pmatrix} = \begin{pmatrix} 0 & \triangle \\ -\triangle & 0 \end{pmatrix} = \tilde{J}
\end{aligned}
$$

(ii) Suppose that the discrete values $x_1, x_2, \ldots, x_N$ determined by the update scheme A are known, and admit the following discrete inverse Fourier transformation

$$x_k = \frac{1}{N}\sum_{n=1}^{N} \tilde{x}_n \exp\left(\frac{2\pi i}{N} kn\right),\ k = 1, \ldots, N.$$

Consider a sequence of discrete points $\{X_k\}_{k=1}^N$ that is shifted by one time step such that $X_k = x_{k+1} = \lambda_1 x_k,\ k = 1, \ldots, N$, i.e., $\{X_k\}_{k=1}^N$ approximates the solution on a later time interval than $\{x_k\}_{k=1}^N$. This admits the following discrete inverse Fourier transformation

$$X_k = \frac{1}{N}\sum_{n=1}^{N} \lambda_1 \tilde{x}_n \exp\left(\frac{2\pi i}{N} kn\right),\ k = 1, \ldots, N,$$

i.e., $\tilde{X}_n = \lambda_1 \tilde{x}_n$. By the definition of the frequency spectrum, it holds $\tilde{X}_n^* \tilde{X}_n = \tilde{x}_n^* \lambda_1^* \lambda_1 \tilde{x}_n = \tilde{x}_n^* |\lambda_1|^2 \tilde{x}_n = \tilde{x}_n^* \tilde{x}_n$, where the last equality relies on the symplecticity. Shifting the discrete solution arbitrary times, we see that the spectrum will be preserved using different time intervals for the frequency analysis. This means that, in particular for long-time integration, a frequency analysis on a later time interval yields the same results as on an earlier time interval, which we denote by *preservation of the frequency spectrum*. The analysis for y follows analogously, and with the linear transformation Q the same holds for q and p. On the other hand, if $|\lambda_{i,j}| \neq 1$, $i, j = 1, 2$ (such as for non-symplectic or non-convergent methods), the frequency spectrum will either shrink or grow unbounded.

Although the analysis was only performed for the simple case of a 1D harmonic oscillator (in particular statement (i) is restricted to this case), we believe that for higher-dimensional systems, a similar statement can also be shown, which is left for future work.

In addition to frequency spectrum, the proposed integrators (e.g., (7.15)) preserve many other structures due to their variational nature, such as symplecticity and momentum maps. Consequently, the correct rate of energy change due to external sources and resistors will be numerically captured, and the sum of all inductor fluxes will be conserved. See [218] for both theoretical theorems and

numerical results.

7.1.5 Noisy circuits

In this section, we extend to simulate noisy circuits, in which noise is added to each branch of the circuit.

Following the description in [40], in the stochastic setting, the constrained stochastic variational principle is

$$\delta \int_0^T \mathcal{L}(q(t), v(t)) + \langle p(t), \dot{q}(t) - v(t) \rangle \, dt + \int_0^T f_L(q(t), v(t), t) \cdot \delta q(t) \, dt + \int_0^T \delta q(t) \cdot (\Sigma \circ dW_t) = 0 \tag{7.17}$$

with constrained variations $\delta q \in \Delta_Q(q)$, where Σ is a $n \times n$ matrix, usually constant and diagonal, indicating the amplitude of noise at each branch, W_t is a n-dimensional Brownian motion, and the last stochastic integral is in the sense of Stratonovich. This principle leads to the constrained stochastic differential equation

$$\frac{\partial \mathcal{L}}{\partial q} - \dot{p} + f_L + \Sigma \circ \frac{dW_t}{dt} \in \Delta_Q^0(q) \tag{7.18a}$$

$$dq = v dt \tag{7.18b}$$

$$\frac{\partial \mathcal{L}}{\partial v} - p = 0 \tag{7.18c}$$

$$K^T v = 0, \tag{7.18d}$$

where by (7.18a) we mean that it holds $\int_0^T \left(\frac{\partial \mathcal{L}}{\partial q} dt - dp + f_L dt + \Sigma \circ dW_t \right) = \int_0^T \mathcal{X}(q) \, dt$ for a vector field $\mathcal{X}(q) \in \Delta_Q^0(q)$ for any T. Correspondingly, the reduced stochastic variational principle reads

$$\delta \int_0^T \mathcal{L}^M(\tilde{q}(t), \tilde{v}(t)) + \left\langle \tilde{p}(t), \dot{\tilde{q}}(t) - \tilde{v}(t) \right\rangle \, dt + \int_0^T f_L^M(\tilde{q}(t), \tilde{v}(t), t) \cdot \delta \tilde{q}(t) \, dt + \int_0^T \delta \tilde{q}(t) \cdot (K_2^T \Sigma \circ dW_t) = 0. \tag{7.19}$$

This results in the reduced stochastic Euler-Lagrange equations

$$\frac{\partial \mathcal{L}^M}{\partial \tilde{q}} dt - d\tilde{p} + f_L^M dt + K_2^T \Sigma \circ dW_t = 0 \tag{7.20a}$$

$$d\tilde{q} = \tilde{v} dt \tag{7.20b}$$

$$\frac{\partial \mathcal{L}^M}{\partial \tilde{v}} - \tilde{p} = 0. \tag{7.20c}$$

To derive the discrete equations with noise, the Stratonovich integral is approximated by a discrete version. For simplicity, we present the equations based on left-point discretization only. On the interval $[t_k, t_{k+1}]$ the integral $\int_{t_k}^{t_{k+1}} \delta\tilde{q}(t) \cdot (K_2^T \Sigma \circ dW_t)$ is approximated by the discrete expression $\delta\tilde{q}_k \cdot (K_2^T \Sigma) B^k$ with $B^k \sim \mathcal{N}(0, h)$, $k = 0, \ldots, N-1$ (see also [40]). In this way, we obtain the following reduced stochastic discrete variational principle

$$\delta \left\{ h \sum_{k=0}^{N-1} \left(\mathcal{L}^M(\tilde{q}_k, \tilde{v}_k) + \left\langle \tilde{p}_k, \frac{\tilde{q}_{k+1} - \tilde{q}_k}{h} - \tilde{v}_k \right\rangle \right) \right\} + h \sum_{k=0}^{N-1} f_L^M(\tilde{q}_k, \tilde{v}_k, t_k) \delta\tilde{q}_k + \sqrt{h} \sum_{k=0}^{N-1} K_2^T \Sigma \xi_k \cdot \delta\tilde{q}_k = 0, \tag{7.21}$$

where for each $k = 0, \ldots, N-1$, ξ_k is a n-dimensional vector with entries being independent standard normal random variables. The discrete reduced stochastic Euler-Lagrange equations that give the symplectic forward Euler iteration scheme is then given by

$$\begin{aligned} \frac{\partial \mathcal{L}^M}{\partial \tilde{q}}(\tilde{q}_k, \tilde{v}_k) - \frac{1}{h}(\tilde{p}_k - \tilde{p}_{k-1}) + f_L^M(\tilde{q}_k, v_k, t_k) + \tfrac{1}{\sqrt{h}} K_2^T \Sigma \xi_k &= 0 \\ \frac{\tilde{q}_k - \tilde{q}_{k-1}}{h} &= \tilde{v}_{k-1} \\ \frac{\partial \mathcal{L}^M}{\partial v}(\tilde{q}_k, \tilde{v}_k) &= \tilde{p}_k \end{aligned}$$

In [40], it is shown that the stochastic flow of a stochastic mechanical system on T^*Q preserves the canonical symplectic form almost surely (i.e., with probability one with respect to the noise). Furthermore, an extension of Noether's theorem

says that in presence of symmetries of the Lagrangian, the corresponding momentum map is preserved almost surely.

7.1.6 Numerical example: High-order LC circuit, stochastic integrator, and multiscale integration

The circuit: Consider a high-order LC circuit given in Figure 7.1. The Kirchhoff Constraint matrix $K \in \mathbb{R}^{n,m}$ and the Fundamental Loop matrix $K_2 \in \mathbb{R}^{n,n-m}$ are (with the third node assumed to be grounded):

$$K = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 0 & -1 \\ -1 & 1 \end{pmatrix}, \qquad K_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \end{pmatrix}. \tag{7.22}$$

Two inductors have inductance $L_1 = 1$ and $L_2 = 1$, and two capacitors have capacitance $C_1 = 1$ and $C_2 = 10$. There are $n = 4$ branches and $m + 1 = 3$ nodes.

With $n_C = 2$ and $K_C = \begin{pmatrix} 0 & -1 \\ -1 & 1 \end{pmatrix}$ having full rank, we can see that the reduced Lagrangian system is non-degenerate, and variational integrator (7.15) can be applied.

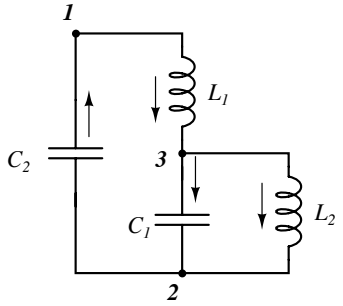


Figure 7.1: Oscillating LC circuit.

Various numerical results in support of the advantage of variational integrators can be found in [218].

Validation on the stochastic variational integrator: A general approach for numerically validating a stochastic variational integrator is the following: consider the stochastic differential equation

$$dx = Axdt + \bar{\Sigma} dW_t, \tag{7.23}$$

where $\bar{\Sigma}$ is a n-by-m matrix, not necessarily full rank, $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, $A \in \mathbb{R}^{n,n}$ and W_t is an m-dimensional Brownian motion (with independent components). The quality of numerical solutions can be evaluated by comparing the empirical statistical moments of the solution to the analytical results. For instance, we can focus on the expectation and the variance, i.e., $\mathbb{E}(x(t))$ and $\mathbb{D}(x(t)) = \mathbb{E}(x(t))^2 - (\mathbb{E}(x(t)))^2$.

On the analytical side, by Ito's formula (see, e.g., [219]) we have with $B(t) = \exp(At)$ and

$$\mathbb{E}(x(t)) = B(t)x(0) \tag{7.24a}$$

$$\mathbb{D}(x(t)) = \int_0^t B(\tau)\bar{\Sigma}\bar{\Sigma}^T B(\tau)^T d\tau. \tag{7.24b}$$

The expectation and the variance can always be computed if A and $\bar{\Sigma}$ are given.

On the numerical side, we run an ensemble of simulations (of total number M), all starting from the same initial condition but for each simulation an independent set of noise (i.e., different ξ_k) is used. The ensemble is indicated by $x^1(t), x^2(t), \ldots, x^M(t)$ where for any j, $x^j(t) = (x_1^j(t), x_2^j(t), \ldots, x_n^j(t))$ is a vector. We compute the empirical moments by

$$\bar{\mathbb{E}}(x(t)) \approx \frac{1}{M}\sum_{j=1}^{M} x^j(t) \tag{7.25a}$$

$$\bar{\mathbb{D}}(x(t)) \approx \frac{1}{M}\sum_{j=1}^{M}(x^j(t))^2 - \frac{1}{M^2}\left(\sum_{j=1}^{M} x^j(t)\right)^2. \tag{7.25b}$$

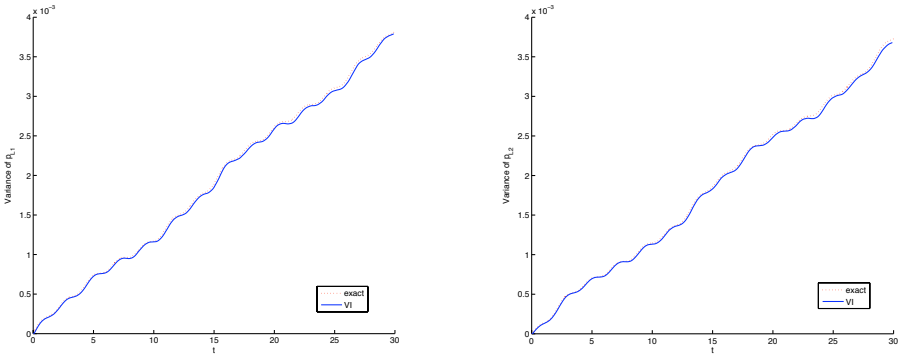
The numerical method is validated if for large enough M the empirical moments (7.25) are close to the analytical ones (7.24).

In our setting, we can rewrite the reduced stochastic Euler-Lagrange equations (7.20) in the form of (7.23) with $x = (\tilde{q}, \tilde{p}) \in \mathbb{R}^{2(n-m)}$, $\bar{\Sigma} = \begin{pmatrix} 0 & 0 \\ 0 & K_2^T\Sigma \end{pmatrix} \in \mathbb{R}^{2(n-m),2n}$, and the obvious definition of $A \in \mathbb{R}^{2(n-m),2(n-m)}$ with $\Sigma \in \mathbb{R}^{n,n}$ and $K_2 \in \mathbb{R}^{n,n-m}$. The analytical variance matrix $\mathbb{D}((\tilde{q}(t), \tilde{p}(t)) \in \mathbb{R}^{2(n-m),2(n-m)}$ for

the reduced system can now be calculated using equation (7.24b). The corresponding variance matrix for the full system can then be calculated as

$$\mathbb{D}(q(t),p(t)) = \begin{pmatrix} K_2 & 0 \\ 0 & K_2 \end{pmatrix} \mathbb{D}(\tilde{q}(t),\tilde{p}(t)) \begin{pmatrix} K_2^T & 0 \\ 0 & K_2^T \end{pmatrix} \in \mathbb{R}^{2n,2n}.$$

As a demonstration, we calculate the empirical and analytical moments for the high-order LC circuit. For the experiments throughout this section, we defined Σ as 4-by-4 diagonal matrix with diagonal entries $\Sigma_{jj} = 0.01$, $j = 1, \ldots 4$. The step size is $h = 0.1$, the integration time for each simulation is $T = 30$, and we start with the initial conditions $\tilde{q}_0 = (1, 0)$, $\tilde{v}_0 = (0, 0)$, and $\tilde{p}_0 = (0, 0)$. The empirical averages are calculated over an ensemble of $M = 100000$ independent simulations.



(a) (b)

Figure 7.2: Benchmark of variances as functions of time according to (7.24b) (red) and variances as functions of time computed numerically by averaging over an ensemble according to (7.25b) (blue). a) $\mathbb{D}p_1$ b) $\mathbb{D}p_2$.

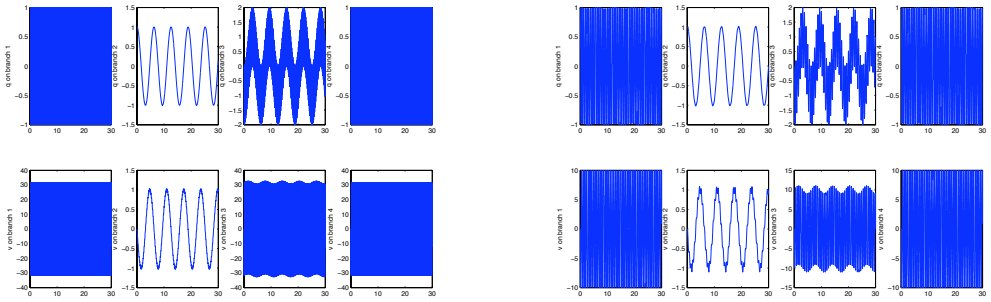
The analytical variance of p_{L_1} and p_{L_2}, i.e., the fifth and sixth diagonal elements of the variance matrix in the full system, are plotted as functions of time (see Figure 7.2, red dotted line). Notice, that p_{L_1} and p_{L_2} in our case are just the currents through inductor branch 1 and 2, the inductances are $L_1 = L_2 = 1$. The result using the stochastic variational integrator is also shown in Figure 7.2 (blue solid line). Both function shapes and ranges agree very well. In particular, all the

little bumps in the variance that are subtly different are approximated correctly. Similarly, empirical expectations well agree with the analytical results too (results not shown).

This classical test serves as evidence that the stochastic integration works well.

7.1.7 Multiscale integration based on FLAVORization

When the circuit exhibits behavior in two time scales, our integrators can be FLAVORized (Chapter 2) to capture the slow time scale without resolving the fast time scale to greatly reduce integration time. For instance, if we regard the capacitance C_2 in the high-order LC circuit as a parameter ϵ, when it has a very small value, there will be a wide separation of timescales, the slow one of which will be strongly captured by FLAVOR.

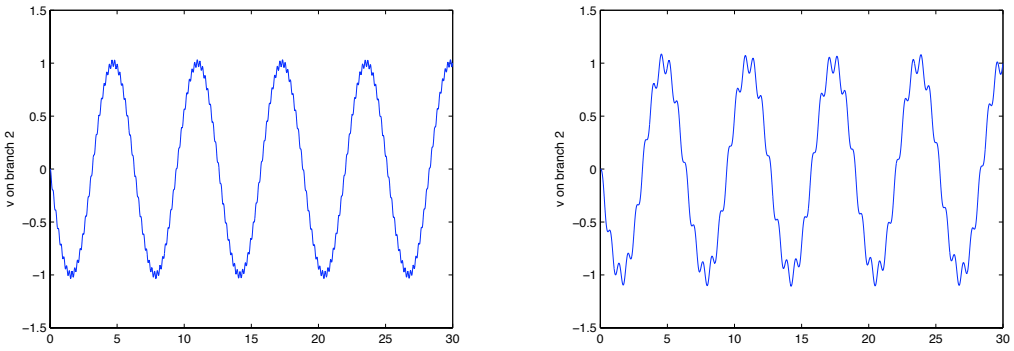


(a) (b)

Figure 7.3: Simulations of a multiscale system: a) Benchmark solution computed with a variational integrator ($h = 10^{-4}$) b) FLAVOR with $\tau = 10^{-4}, \delta = 10^{-3}$ and $\epsilon = 10^{-3}$.

Specifically, FLAVORize our variational circuit integator (7.15) by the rule (2.8). $C_2 = \epsilon = 10^{-3}$, $\tau = 0.1\epsilon = 10^{-4}$, $H = 0.1$ and $M = 100$. The charges and currents as functions of time are plotted in Figure 7.3. Notice that the slow components in the solution are captured strongly, but the fast components may have altered wave shapes: for instance, Figure 7.4 shows a zoomed-in investigation of the current through the second branch, which is a superposition of a slow global

oscillation and a fast local oscillation; the slow one is obviously well-captured in the usual sense, and the fast one is captured in the less-commonly-used sense of averaging.



(a) (b)

Figure 7.4: Simulations of a multiscale system: a) Benchmark solution computed with a variational integrator ($h = 10^{-4}$) b) FLAVOR with $\tau = 10^{-4}, \delta = 10^{-3}$ and $\epsilon = 10^{-3}$.

7.2 Frequency domain method for nonlinear wave propagation

By going to the frequency domain, the propagation of acoustic wave in a nonlinear homogeneous medium (originally modeled by a nonlinear wave equation) could be represented by a system of ODEs. These ODEs could be efficiently integrated by techniques analogous to those used by the impulse methods (Chapter 4).

Many results in this section are excerpts or paraphrases of the content of a published paper [160]. Only the part on numerical integration will be included in this thesis, and the original paper is referred to for modeling and acoustical applications.

7.2.1 The formulation in frequency domain

Consider time-domain Westervelt equation in a homogeneous medium

$$\nabla^2 p(\mathbf{r},t) - \frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) + \frac{\delta}{c_0^4}\frac{\partial^3}{\partial t^3}p(\mathbf{r},t) + \frac{\beta}{\rho_0 c_0^4}\frac{\partial^2}{\partial t^2}p^2(\mathbf{r},t) = 0, \tag{7.26}$$

where $\mathbf{r} \in \mathbb{R}^3$ is the spatial variable, $t \in \mathbb{R}$ is the temporal variable, p is the sound pressure, c_0 is the sound speed, δ is the sound diffusivity, β is the nonlinearity coefficient, and ρ_0 is the ambient density.

By Fourier transforming the temporal dimension as well as the Cartesian $x-$ and $y-$dimensions, we obtain a system of ODEs:

$$\frac{\partial^2}{\partial z^2}P(k_x,k_y,z,\omega) + K^2 P(k_x,k_y,z,\omega) - \frac{\beta\omega^2}{\rho_0 c_0^4}P(k_x,k_y,z,\omega)\otimes P(k_x,k_y,z,\omega) = 0 \tag{7.27}$$

where

$$P(k_x,k_y,z,\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y,z,t)\exp(-i(k_x x + k_y y - \omega t)\,dx\,dy\,dt, \tag{7.28}$$

$$P(k_x,k_y,z,\omega)\otimes P(k_x,k_y,z,\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(k_x',k_y',z,\omega')P(k_x-k_x',k_y-k_y',z,\omega-\omega')\,dk_x'\,dk_y'\,d\omega', \tag{7.29}$$

$$K^2 = \frac{\omega^2}{c_0^2} - k_x^2 - k_y^2 - \frac{\delta\omega^3}{c_0^4} \tag{7.30}$$

Define $M = \frac{\beta\omega^2}{\rho_0 c_0^4}$ and a nonlinear force to be

$$F(P(z')) := P(k_x,k_y,z',\omega)\otimes P(k_x,k_y,z',\omega), \tag{7.31}$$

and schematically suppress k_x, k_y and ω, then (7.27) can be rewritten as

$$\frac{\partial^2}{\partial z^2}P(z) + K^2 P(z) - MF(P(z)) = 0 \tag{7.32}$$

Since we are interested in how the wave propagates from $z = 0$, this is a well defined initial value problem: for each z, obviously $K^2P(z) - MF(P(z))$ is known, and P at a larger z can be thenceforth obtained.

7.2.2 Integration by uniform macroscopic steps

Due to underlying acoustic reasons, in the region of interests, M is always $\mathcal{O}(1)$, but K could be $\gg 1$. Since the nonlinear force (based on a convolution) involves all degrees of freedom (indexed by k_x, k_y and ω), to numerically integrate (7.27), a uniform integration step (on z) is needed. Consequently, constrained by the stiffest K, a single-scale ODE solver requires a microscopic step on z.

However, since (7.27) is nothing but a forced mechanical system with a quadratic stiff potential (this could be better seen in the form of (7.32)), we could use the idea explained in Chapter 4 to numerically integrate the system using a macroscopic step Δz (independent of the stiff K):

Following the language of [160], the solution to (7.32) is

$$P(z) = P(0)e^{iKz} + \frac{Me^{iKz}}{2iK}\int_0^z e^{-iKz'}F(P(z'))\,dz' \tag{7.33}$$

To obtain its numerical solution, approximate the integral by using the propagator of the stiff linear force:

$$\begin{aligned}
&\int_0^{\Delta z} e^{-iKz'}F(P(z'))\,dz' \approx F(P(0))\Delta z \\
&\int_0^{2\Delta z} e^{-iKz'}F(P(z'))\,dz' \approx F(P(0))\Delta z + e^{-iK\Delta z}F(P(\Delta z))\Delta z \\
&\cdots
\end{aligned} \tag{7.34}$$

Naturally, this is equivalent to a 1st-order impulse method (see Remark 4.2.1). As a

consequence, we have both guaranteed accuracy and accelerated the computation. Again, [160] is referred to for numerical illustrations, as well as this method's acoustic implications. Good results are, of course, obtained.

7.3 Optimization of Freidlin-Wentzell theory and mass effect

In a SDE, the probability of transiting from one state to another state could be characterized by Freidlin-Wentzell large deviation theory [107] in the weak noise limit. By optimizing the transition rate functional in a trajectory space, the optimal transition pathway between two states could be obtained.

Two approaches for this optimization in a Langevin setting are presented in this section, the first is an analytical method that works for any two states in a linear system, and the second is based on the time reparameterization of the transition pathway given by an inertial version of the String method [90, 91], which uses a gradient algorithm to compute the path between two metastable states in an arbitrary potential landscape.

This study is motivated by the observation of mass effects in molecular dynamics [300], and we show that significant mass effects can be quantified by different rates of optimal transitions for different masses. This also suggests that the use of overdamped Langevin $c\,dq = -\nabla V(q)\,dt - dW_t$ in molecular dynamics is not always justified.

7.3.1 Rate functional for Langevin equations

Being a large deviation theory [80], Freidlin-Wentzell theory [107] works in path space as follows:

Given a stochastic dynamical system

$$d\boldsymbol{X}_\epsilon(t) = \boldsymbol{b}(\boldsymbol{X}_\epsilon)dt + \epsilon^{1/2}\boldsymbol{\sigma}(\boldsymbol{X}_\epsilon)d\boldsymbol{W}(t), \tag{7.35}$$

there is a rate functional over $\mathcal{C}^n[0,T]$ defined as follows:

$$I(\boldsymbol{\phi}) = \int_0^T J(\boldsymbol{\phi}(t), \dot{\boldsymbol{\phi}}(t))dt, \tag{7.36}$$

where $\boldsymbol{\phi}(0) = \boldsymbol{x}$, and assuming diffusion matrix $\boldsymbol{A} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T$ is uniformly positive definite,

$$J(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{b}(\boldsymbol{x}))^T \boldsymbol{A}^{-1}(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{b}(\boldsymbol{x})) \tag{7.37}$$

The rate functional describes the asymptotic behavior of large deviation, in the sense that, given $\boldsymbol{X}_\epsilon(t)$ being the solution to (7.35) with initial condition $\boldsymbol{X}_\epsilon(0) = \boldsymbol{x}$, we have

$$P(\sup_{0\le t\le T} |\boldsymbol{X}_\epsilon(t) - \boldsymbol{\phi}(t)| < \delta) \sim \exp(-\epsilon^{-1} I(\boldsymbol{\phi})), \epsilon \to 0 \tag{7.38}$$

for any small $\delta > 0$.

Therefore, one seeks for a $\boldsymbol{\phi}(t)$ in certain path space which minimizes the rate functional as the most probable path. If one is interested in transition between configurations, the path space could be $\{\phi \in \mathcal{C}[0,T] | \phi(0) = \boldsymbol{A}, \phi(T) = \boldsymbol{B}\}$ or $\bigcup_{T>0}\{\phi \in \mathcal{C}[0,T] | \phi(0) = \boldsymbol{A}, \phi(T) = \boldsymbol{B}\}$.

When the system is Langevin, in which the noise is degenerate, i.e.,

$$\begin{cases} d\boldsymbol{q} &= \boldsymbol{M}^{-1}\boldsymbol{p}\, dt \\ d\boldsymbol{p} &= -\nabla V(\boldsymbol{q})\, dt - c\boldsymbol{p}\, dt + \epsilon^{1/2}\boldsymbol{\sigma}\, dW \end{cases}, \tag{7.39}$$

where the temperature T of this system satisfies $\frac{2c}{\epsilon\sigma^2} = \frac{1}{T}$, the corresponding integrand of the rate functional I could be shown by large deviation theory as

$$\begin{cases} J(\boldsymbol{q}, \boldsymbol{p}) &= \frac{1}{2}(\dot{\boldsymbol{p}} + c\boldsymbol{p} + \nabla V(\boldsymbol{q}))^2 \qquad \text{if } \boldsymbol{p} = \boldsymbol{M}\dot{\boldsymbol{q}} \\ &= \infty \qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{7.40}$$

Therefore we can study the following constrained variational problem for the most probable transition from A to B:

$$\begin{cases} \delta I & = 0 \\ I & = \int_0^T \frac{1}{2}(\dot{\boldsymbol{p}} + c\boldsymbol{p} + \nabla V(\boldsymbol{q}))^2 dt \\ \boldsymbol{p} & = \boldsymbol{M}\dot{\boldsymbol{q}} \\ \boldsymbol{q}(0) & = A, \\ \boldsymbol{q}(T) & = B \end{cases}$$

In molecular dynamics, one is actually more interested in an alternative version which takes the Gibbs-Boltzmann distribution of kinetic energy into account, because otherwise the solution will always be a Newtonian path with a big enough initial velocity to overcome all energy barrier along its way, and this path will render $I(\cdot)$ zero. Therefore, instead of minimizing I, we minimize $\mathcal{A} = \frac{I}{2ckT} + \frac{\boldsymbol{p}(0)^T \boldsymbol{M}^{-1} \boldsymbol{p}(0)}{2kT}$ under same constraints. Notice that the probability will change as the temperature changes, but the optimal path will not.

7.3.2 An analytical solver

The approach is: first, fix $\boldsymbol{p}(0)$ and solve the variational problem without the end point constraint $q(T) = B$; then, optimize among solutions that satisfy $q(T) = B$; finally, optimize with respect to $\boldsymbol{p}(0)$. More precisely, introduce a Lagrange multiplier λ on the cotangent space and use Hamilton-Pontryagin principle:

$$0 = \delta \int_0^T \frac{1}{2}\|\dot{p} + \nabla V(q) + cp\|_2^2 + \lambda(p - M\dot{q})dt \tag{7.41}$$

Taking the variation leads to the following ODE system, whose solution is the solution to the variational problem:

$$\begin{cases} -(\ddot{p} + c\dot{p} + \nabla\nabla V(q)\dot{q}) + (\dot{p} + cp + \nabla V(q))c + \lambda & = 0 \\ (\dot{p} + cp + \nabla V(q))\nabla\nabla V(q) + M\dot{\lambda} & = 0 \\ p & = M\dot{q} \end{cases} \tag{7.42}$$

One set of sufficient initial conditions is $q(0), \dot{q}(0), \ddot{q}(0), \dddot{q}(0)$; $q(t)$ will be a

function of (only) them. $q(0) = A$ is known, and we first assume $\dot{q}(0)$ is known and fixed as well. Then, we optimize I with respect to values of $\ddot{q}(0)$ and $\dddot{q}(0)$ under the constraint of $q(T) = B$, and the optimal I is a function of $\dot{q}(0)$. Finally, we allow $\dot{q}(0)$ to change and minimize $\mathcal{A}$ as a function of $\dot{q}(0)$. This way, the optimal transition path, represented by $\dot{q}(0), \ddot{q}(0), \dddot{q}(0)$ and (7.42), is obtained.

When the potential is quadratic, the ODE system (7.42) is linear, and an exact solution exists. In this situation, the constrained minimization could be solved analytically. Details are omitted.

When the potential is non-quadratic, the entire procedure described above is still valid, but the satisfaction of $q(T) = B$ is a shooting problem that is numerically difficult to solve.

7.3.3 A numerical solver

The numerical solver contains two steps; the first step is known and could be extracted from the literature on the String method, and the second step is our new contribution.

String method for inertial Langevin: One important observation that the String method [90, 91] made use of is that the optimal transition path between two minima of $V(\cdot)$ in a system

$$dx = -\nabla V(x)\, dt + \sigma\, dW_t \tag{7.43}$$

must satisfy

$$(\nabla V)^{\perp}(x) = 0, \tag{7.44}$$

where $(\nabla V)^{\perp}(x)$ indicates for each t the vector value corresponding to ∇V projected in the direction perpendicular to the tangent of the path x (i.e., $\dot{x}$).

This observation could be generalized to systems of the form (see [189, 190] for similar examples):

$$dx = K\nabla H(x)\, dt + \Sigma\, dW_t, \tag{7.45}$$

and the optimal path will analogously satisfy

$$(K\nabla H)^{\perp}(x) = 0 \tag{7.46}$$

Therefore, the String method (with a small modification described below) will work for Langevin equations, because under the notation $x = (q, p)$, $K = \begin{bmatrix} 0 & I \\ -I & -c \end{bmatrix}$, $H(x) = V(q) + \frac{1}{2}p^T M^{-1} p$ and $\Sigma = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}$, (7.45) rewrites to be Langevin:

$$\begin{cases} dq &= M^{-1}dp \\ dp &= -\nabla V(q)\,dt - cp\,dt + \sigma\,dW_t \end{cases} \tag{7.47}$$

Recall that the String method is an iteration of two half steps, the first one being an evolution of sample points on the string, and the second being a resampling, which ensures that sample points are uniformly distributed on the string [91]. Without quoting further details of the String method, we just point out that the modification for (7.45) is to use $K\nabla H$ instead of ∇H as the drift in the evolution (i.e., in the first half step), and the second half step remains unchanged.

Time reparameterization: (7.46) is a necessary condition for the minimization of the rate functional (7.36) (with integrand (7.40)), but not sufficient. This is because there can be multiple paths with different time parameterizations that satisfy (7.46). In fact, the path produced by the inertial String method (the previous step) is most likely not the minimizer. Therefore, we propose the following (a second step) to find the minimizer:

Having String method's result at hand, we further minimize I (or $\mathcal{A}$ in the context of molecular dynamics) with respect to a time reparameterization. More precisely, suppose the String method yields a path $x(\tau), \tau \in [0, L]$, we look for an increasing function $\mathcal{T} : t \mapsto \tau$ with $\mathcal{T}(0) = 0$ and $\mathcal{T}(T) = L$, such that $I(x(\mathcal{T}(\cdot)))$ (or $\mathcal{A}(x(\mathcal{T}(\cdot)))$) is optimized (with respect to all possible $\mathcal{T}$. This

optimization is easy because $\mathcal{T}$ is 1-dimensional ($[0,T] \to [0,L]$), and the objective function depends on the reparameterization in a weakly nonlinear fashion (the reparameterization does not enter q, hence not $V(q)$).

For numerical implementations, piece-wise linear interpolations are always used.

7.3.4 A molecular example of mass effect

As an example to illustrate mass effects in molecular dynamics, consider the Langevin dynamics of three atoms with a pair-wise Morse potential, i.e., (7.47) with

$$V(q) = e^{-2(d_{12}-d_0)} - 2e^{-(d_{12}-d_0)} + e^{-2(d_{13}-d_0)} - 2e^{-(d_{13}-d_0)} + e^{-2(d_{23}-d_0)} - 2e^{-(d_{23}-d_0)}, \tag{7.48}$$

where $q \in \mathbb{R}^6$ represents the positions of three atoms in a plane, and $d_{ij} = \sqrt{(q_{2i-1}-q_{2j-1})^2 + (q_{2i}-q_{2j})^2}$ is the pair-wise distance between atoms.

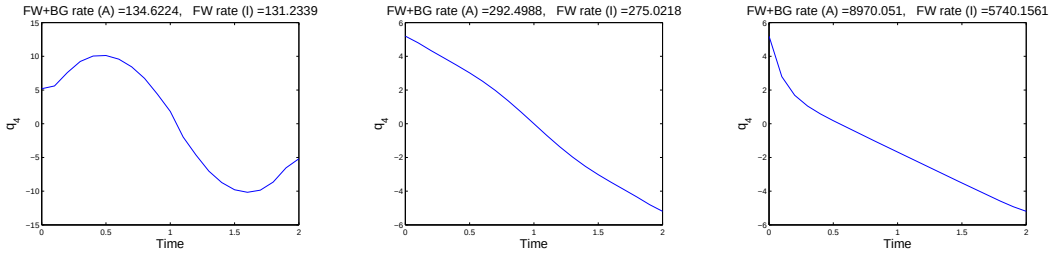


Figure 7.5: Optimal transition paths illustrated by q_4 (y-coordinate of the second atom) for different masses of the second atom. The second atom mass is respectively 0.1 (left), 1 (middle) and 10 (right), and the other two masses are both 1 and always fixed. The friction coefficient $c = 10$, and the temperature $\frac{\sigma^2}{2c} = 1$, both fixed. Total transition time $T = 2$ is also fixed. More technical parameters are: the number of sample points on the string $N = 100$, the string evolution timestep $h = 0.01$, and the time reparameterization interpolation step size $\tilde{h} = 0.1$.

Consider $d_0 = 6$ and the optimal transition from $q(0) = [-3, -\sqrt{3}, 0, 2\sqrt{3}, 3, -\sqrt{3}]$ to $q(T) = [-3, \sqrt{3}, 0, -2\sqrt{3}, 3, \sqrt{3}]$, both of which are local minima of the potential. The transition between these two metastable states corresponds to a reaction of isomerization, and a movie of this reaction is available at `http://www.youtube.com/watch?v=MqZ1_t9z1Uw` .

Figure 7.5 compares the optimal transition paths for different mass values of the second atom. Results are computed by the numerical solver proposed in Section 7.3.3 (the analytical solver in Section 7.3.2 does not apply here, and its results, when applicable, are less interesting). Significant difference can be seen, not only on the trajectory, but also on the associated rate (which is negative of the reaction rate; hence, a bigger value indicates a less possible reaction). The results are very natural, because it is more difficult for a heavier mass to move.

This provides a probabilistic confirmation of the mass effects in this system, which was illustrated numerically and explained in a framework of differential geometry in [300].

Of course, other parameters in the system, such as the temperature or the friction coefficient, affect the transition significantly as well. The investigation on their effects will be analogous.

Chapter 8

Future directions

Along the line of multiscale integration, there are at least three unsolved problems. The first is the case in which the scale separation is due to nonlinearity but not stiffness. An illustrative example is a Hamiltonian system with $\mathcal{H} = p_1^2/2 + p_2^2/2 + q_1^2/2 + (q_2 - q_1)^{10}$, where in the nonlinear regime $q_2 - q_1$ acts like a stiff spring. A more realistic example is shown in [301], where the dynamics of a cluster of inert atoms with pair-wise Morse potential (which models van der Waals forces) exhibits clear separation of timescales. Similar phenomena will happen in a lot of complicated systems, such as proteins. One possible approach to this problem is to use an extension of artificial FLAVORs (Section 2.2.6), in which the nonlinearity is frozen over mesoscopic (with size $\delta - \tau$) substeps.

The second problem is how to treat a broad yet not well-separated spectrum of timescales (if adjacent timescales are well-separated, FLAVORs can be used in a nested way). Protein dynamics will be a typical example, in which bond oscillations have a characteristic frequency around 10^{14}Hz, bond-angle oscillations at 10^{13}Hz, torsion dynamics at 10^{12}Hz, and non-covalence effects (due to van der Waals forces, electrostatic forces, and so on) are even slower. The investigation of this problem possibly requires a significant extension of the mathematical notion of ergodic measure.

The third open question is, can numerical methods respect the symmetry between the space and time in multiscale PDEs? At least according to numerical

experiments, PDE-FLAVORs (Chapter 3) are able to handle stiffness in both space and time, but its current error analysis relies on introductions of semi-discrete systems, in which the space is discrete but the time is continuous, and slow and fast variables are both defined in terms of the semi-discrete systems. Whether the integrator itself intrinsically breaks the space-time symmetry is not clear at this stage, and it is possible that a more symmetric proof is yet to be found. One possibility is to look for a generalized averaging theorem (analogous to (2.5)) using a local ergodic measure on fields ([32] might be a relevant reference).

At the same time, possible interplays between different methods proposed in this thesis should be explored. For instance, the temperature and friction accelerated sampling approach (Chapter 6) could employ a FLAVOR (Chapter 2) as its base Langevin integrator. Whether this results in an ergodic scheme is worth a mathematical investigation, and if it does, the next question would be on the form of the corresponding ergodic measure. By answering these two questions, we might be able to justify the multiscale nature of the resulting sampling algorithm. Another possibility is to combine FLAVORs (Chapter 2) with Freidlin-Wentzell optimizers (Section 7.3) to probe optimal transition pathways in multiscale systems. A third potential subject would be to propose multiscale geometric integrators for stochastic partial differential equations (SPDEs).

Moreover, many interesting applications are unexplored. We will not enumerate direct applications, such as important multiscale systems that our generic integrators could simulate. Instead, we wish to point out that the analytical results in Section 4.3 provide a way to take the derivative of a parameter-dependent matrix exponential. This way, the result will be in a closed-form, which is a significant improvement from the current theory based on Magnus expansion, which is a sum of an infinite series of matrix commutators [187].

A relevant topic is the phenomenon of parametric resonance. A classical example is the system:

$$\ddot{x} + \beta(t)\dot{x} + \omega^2 x = 0 \tag{8.1}$$

In this system, if $\beta(t)$ is chosen to be $\cos(2\omega t)$ or $\sin(2\omega t)$, the oscillation amplitude will grow exponentially in time. The mathematical theory (especially in the case of nonlinear systems) of this phenomenon needs development, and its real-life applications need to be carefully designed.

To put parametric resonance in a more general framework, an interesting future direction would be the temporal homogenization of (controlled) mechanical systems. Classical literature that stimulates this topic could date back to Mathieu's equation [196], Hill's equation [142], Floquet theory [105], etc.

Appendix A

Appendix: Additional proofs

A.1 Proof of Theorems 2.1.1 and 2.1.2

Define the process $t \mapsto (\bar{x}_t, \bar{y}_t)$ by

$$(\bar{x}_t, \bar{y}_t) := \eta(\bar{u}_t). \tag{A.1}$$

It follows from the regularity of η that it is sufficient to prove the F-convergence of $(\bar{x}_t, \bar{y}_t)$ towards $\delta_{X_t} \otimes \mu(X_t, dy)$. Moreover, it is also sufficient to prove inequalities (A.2) and (A.3) in order to obtain inequalities (2.24) and (2.25)

$$|x_t^\epsilon - \bar{x}_t| \leq Ce^{Ct}\psi_1(u_0, \epsilon, \delta, \tau) \tag{A.2}$$

and

$$\Big|\frac{1}{T}\int_t^{t+T} \varphi(\bar{x}_s, \bar{y}_s)\, ds - \int_{\mathbb{R}^p} \varphi(X_t, y)\mu(X_t, dy)\Big| \leq \psi_2(u_0, \epsilon, \delta, \tau, T, t)(\|\varphi\|_{L^\infty} + \|\nabla\varphi\|_{L^\infty}) \tag{A.3}$$

Now define ψ_τ^ϵ by

$$\psi_\tau^\epsilon(x, y) := \eta \circ \theta_\tau^\epsilon \circ \eta^{-1}(x, y) \tag{A.4}$$

Define ψ_h^g by

$$\psi_h^g(x, y) := \eta \circ \theta_h^G \circ \eta^{-1}(x, y) \tag{A.5}$$

Proposition A.1.1. *The vector fields f and g associated with the system* (2.2) *are Lipschitz continuous. We also have*

$$(\bar{x}_t, \bar{y}_t) = \big(\psi^g_{\delta-\tau} \circ \psi^\epsilon_\tau\big)^k (x_0, y_0) \quad \textit{for} \quad k\delta \le t < (k+1)\delta. \tag{A.6}$$

Moreover, there exists $C > 0$ such that for $h \le h_0$ and $\frac{\tau}{\epsilon} \le \tau_0$ we have

$$\big|\psi^\epsilon_\tau(x,y) - (x,y) - \tau\big(g(x,y),0\big) - \frac{\tau}{\epsilon}\big(0, f(x,y)\big)\big| \le C\big(\frac{\tau}{\epsilon}\big)^2 \tag{A.7}$$

and

$$\big|\psi^g_h(x,y) - (x,y) - h\big(g(x,y),0\big)\big| \le Ch^2. \tag{A.8}$$

Furthermore, given x_0, y_0, the trajectories of $(x^\epsilon_t, y^\epsilon_t)$ and $(\bar{x}_t, \bar{y}_t)$ are uniformly bounded in ϵ, $\delta \le h_0$, $\tau \le \min(\tau_0\epsilon, \delta)$.

Proof. Since $(x,y) = \eta(u)$, we have

$$\dot{x} = (G + \frac{1}{\epsilon}F)\nabla\eta^x \circ \eta^{-1}(x,y) \tag{A.9}$$

$$\dot{y} = (G + \frac{1}{\epsilon}F)\nabla\eta^y \circ \eta^{-1}(x,y). \tag{A.10}$$

Hence, we deduce from (2.2) in Condition 2.1.1 that

$$g(x,y) = G\nabla\eta^x \circ \eta^{-1}(x,y) \tag{A.11}$$

$$f(x,y) = F\nabla\eta^y \circ \eta^{-1}(x,y). \tag{A.12}$$

We deduce the regularity of f and g from the regularity of G, F and η. (A.6) is a direct consequence of the definition of ψ^ϵ_τ and ψ^g_h and (2.27) (we write $(x_0, y_0) := \eta(u_0)$). Observe that (2.2) in Condition 2.1.1 also requires that

$$F\nabla\eta^x = 0, \qquad G\nabla\eta^y = 0. \tag{A.13}$$

Now observe that

$$\psi^\epsilon_\tau(x,y) - (x,y) - \big(g(x,y),0\big)\tau - \big(0,f(x,y)\big)\frac{\tau}{\epsilon} = \\ \big(\eta\circ\theta^\epsilon_\tau - \eta - \tau\big(G\nabla\eta^x,0\big) - \frac{\tau}{\epsilon}\big(0,F\nabla\eta^y\big)\big)\circ\eta^{-1}(x,y). \tag{A.14}$$

Using (A.13), (2.29), Taylor expansion, and the regularity of η, we obtain (A.7). Similarly

$$\psi^g_h(x,y)-(x,y)-h\big(g(x,y),0\big) := \Big(\eta\circ\theta^G_h-\eta(x,y)-h\big(G\nabla\eta^x,0\big)\Big)\circ\eta^{-1}(x,y). \tag{A.15}$$

Using (A.13), (2.28), Taylor expansion and the regularity of η we obtain (A.8). The uniform bound (depending on x_0, y_0) on the trajectories of $(x^\epsilon_t, y^\epsilon_t)$ and $(\bar{x}_t, \bar{y}_t)$ is a consequence of the uniform bound (given u_0) on the trajectories of u^ϵ_t and $\bar{u}_t$. □

It follows from Proposition A.1.1 that it is sufficient to prove Theorems 2.1.1 and 2.1.2 in the situation where η is the identity diffeomorphism. More precisely, the F-convergence of $\bar{u}_t$ is a consequence of the F-convergence of $(\bar{x}_t, \bar{y}_t)$ and the regularity of η. Furthermore, from the uniform bound (depending on (x_0, y_0)) on the trajectories of $(x^\epsilon_t, y^\epsilon_t)$ and $(\bar{x}_t, \bar{y}_t)$ we deduce that g and f are uniformly bounded and Lipschitz continuous (in ϵ, $\delta \leq h_0$, $\tau \leq \min(\tau_0\epsilon, \delta)$) over those trajectories.

Define

$$\bar{g} := \int g(x,y)\,\mu(x,dy)$$

where μ is the family of measures introduced in Condition 2.1.2. Let us prove the following lemma:

Lemma A.1.1.

$$|x^\epsilon_{n\delta} - \bar{x}_{n\delta}| \leq Ce^{Cn\delta}\Big(\delta + \big(\frac{\tau}{\epsilon}\big)^2\frac{1}{\delta} + \sup_{1\leq l\leq n}|J(l)|\Big) \tag{A.16}$$

with $J(k) = J_1(k) + J_2(k)$,

$$J_1(k) := \sum_{n=0}^{k-1} \Big(\int_{n\delta}^{(n+1)\delta} g(x^\epsilon_{n\delta}, y^\epsilon_s)\, ds - \delta \bar{g}(x^\epsilon_{n\delta}) \Big) \tag{A.17}$$

and

$$J_2(k) := \sum_{n=0}^{k-1} \delta \big(\bar{g}(\bar{x}_{n\delta}) - g(\bar{x}_{n\delta}, \bar{y}_{n\delta}) \big) \tag{A.18}$$

Proof. Observe that

$$x^\epsilon_{(n+1)\delta} = x^\epsilon_{n\delta} + \int_{n\delta}^{(n+1)\delta} g(x^\epsilon_{n\delta}, y^\epsilon_s)\, ds + \int_{n\delta}^{(n+1)\delta} (g(x^\epsilon_s, y^\epsilon_s) - g(x^\epsilon_{n\delta}, y^\epsilon_s))\, ds \tag{A.19}$$

Hence,

$$x^\epsilon_{(n+1)\delta} - \bar{x}_{(n+1)\delta} = x^\epsilon_{n\delta} - \bar{x}_{n\delta} + I_1 + I_2(n) + I_3 + I_4(n) + I_5 \tag{A.20}$$

with

$$I_1 := \int_{n\delta}^{(n+1)\delta} (g(x^\epsilon_s, y^\epsilon_s) - g(x^\epsilon_{n\delta}, y^\epsilon_s)) ds \tag{A.21}$$

$$I_2(n) := \int_{n\delta}^{(n+1)\delta} g(x^\epsilon_{n\delta}, y^\epsilon_s)\, ds - \delta \bar{g}(x^\epsilon_{n\delta}) \tag{A.22}$$

$$I_3 := \delta \big(\bar{g}(x^\epsilon_{n\delta}) - \bar{g}(\bar{x}_{n\delta}) \big) \tag{A.23}$$

$$I_4(n) := \delta \big(\bar{g}(\bar{x}_{n\delta}) - g(\bar{x}_{n\delta}, \bar{y}_{n\delta}) \big) \tag{A.24}$$

$$I_5 := \delta g(\bar{x}_{n\delta}, \bar{y}_{n\delta}) - (\bar{x}_{(n+1)\delta} - \bar{x}_{n\delta}) \tag{A.25}$$

Now observe that

$$|I_1| \le \|\nabla_x g\|_{L^\infty} \delta^2 \tag{A.26}$$

and

$$|I_3| \le \delta \|\nabla_x g\|_{L^\infty} |x^\epsilon_{n\delta} - \bar{x}_{n\delta}|. \tag{A.27}$$

Using (A.7) and (A.8) we obtain that

$$|I_5| \leq C\Big(\delta^2 + \big(\frac{\tau}{\epsilon}\big)^2\Big) \tag{A.28}$$

Combining the previous equations, we obtain

$$x^\epsilon_{(n+1)\delta} - \bar{x}_{(n+1)\delta} \leq x^\epsilon_{n\delta} - \bar{x}_{n\delta} + C\Big(\delta^2 + \big(\frac{\tau}{\epsilon}\big)^2\Big) + C\delta|x^\epsilon_{n\delta} - \bar{x}_{n\delta}| + (I_2 + I_4)(n) \tag{A.29}$$

and

$$x^\epsilon_{(n+1)\delta} - \bar{x}_{(n+1)\delta} \geq x^\epsilon_{n\delta} - \bar{x}_{n\delta} - C\Big(\delta^2 + \big(\frac{\tau}{\epsilon}\big)^2\Big) - C\delta|x^\epsilon_{n\delta} - \bar{x}_{n\delta}| + (I_2 + I_4)(n) \tag{A.30}$$

Write

$$J(n) := \sum_{k=0}^{n-1} (I_2 + I_4)(k) \tag{A.31}$$

Summing the first n inequalities (A.29) and (A.30), we obtain

$$x^\epsilon_{n\delta} - \bar{x}_{n\delta} \leq C\Big(\delta^2 + \big(\frac{\tau}{\epsilon}\big)^2\Big)n + C\delta \sum_{k=0}^{n-1} |x^\epsilon_{k\delta} - \bar{x}_{k\delta}| + J(n) \tag{A.32}$$

$$x^\epsilon_{n\delta} - \bar{x}_{n\delta} \geq -C\Big(\delta^2 + \big(\frac{\tau}{\epsilon}\big)^2\Big)n - C\delta \sum_{k=0}^{n-1} |x^\epsilon_{k\delta} - \bar{x}_{k\delta}| + J(n) \tag{A.33}$$

Hence

$$|x^\epsilon_{n\delta} - \bar{x}_{n\delta}| \leq C\Big(\delta^2 + \big(\frac{\tau}{\epsilon}\big)^2\Big)n + C\delta \sum_{k=0}^{n-1} |x^\epsilon_{k\delta} - \bar{x}_{k\delta}| + |J(n)| \tag{A.34}$$

And we obtain by induction

$$\begin{aligned}|x^\epsilon_{n\delta} - \bar{x}_{n\delta}| \leq & C\Big(\delta^2 + \big(\frac{\tau}{\epsilon}\big)^2\Big)\big(n + C\delta\sum_{k=1}^{n}(n-k)(1+C\delta)^{k-1}\big) \\ & + |J(n)| + C\delta\sum_{l=2}^{n}(1+C\delta)^{l-2}|J(n-l+1)|\end{aligned} \tag{A.35}$$

Equation (A.35) concludes the proof of Lemma A.1.1. □

We now need to control $J_1(k)$ and $J_2(k)$. First, let us prove the following lemma.

Lemma A.1.2. *For $N \in \mathbb{N}^*$, we have*

$$|J_1(k)| \leq (\delta k)C\Big(\delta e^{C\frac{\delta}{N\epsilon}} + E\big(\frac{\delta}{N\epsilon}\big)\Big) \tag{A.36}$$

Proof. Define $\hat{y}^\epsilon_t$ such that $\hat{y}^\epsilon_t = y^\epsilon_t$ for $t = (n + j/N)\delta$, $j \in \mathbb{N}^*$, and

$$\frac{d\hat{y}^\epsilon_t}{dt} = \frac{1}{\epsilon}f(x^\epsilon_{n\delta}, \hat{y}^\epsilon_t) \quad \text{for} \quad (n+j/N)\delta \leq t < (n+(j+1)/N)\delta. \tag{A.37}$$

Using the regularity of f and g, we obtain

$$|\hat{y}^\epsilon_t - y^\epsilon_t| \leq C\delta e^{C\frac{\delta}{N\epsilon}}. \tag{A.38}$$

First, observe that

$$\frac{1}{\delta}\int_{n\delta}^{(n+1)\delta} g(x^\epsilon_{n\delta}, y^\epsilon_s)\,ds - \bar{g}(x^\epsilon_{n\delta}) = K_1 + K_2 \tag{A.39}$$

with

$$K_1 := \frac{1}{\delta}\sum_{j=0}^{N-1}\int_{(n+j/N)\delta}^{(n+(j+1)/N)\delta}\big(g(x^\epsilon_{n\delta}, y^\epsilon_s) - g(x^\epsilon_{n\delta}, \hat{y}^\epsilon_s)\big)\,ds \tag{A.40}$$

and

$$K_2 := \frac{1}{N}\sum_{j=0}^{N-1}\Big(\frac{N}{\delta}\int_{(n+j/N)\delta}^{(n+(j+1)/N)\delta} g(x^\epsilon_{n\delta},\hat{y}^\epsilon_s)\,ds - \bar{g}(x^\epsilon_{n\delta})\Big). \tag{A.41}$$

We have

$$|K_1| \le \|\nabla_y g\|_{L^\infty}\frac{1}{N}\sum_{j=0}^{N-1}\sup_{(n+j/N)\delta\le s\le(n+(j+1)/N)\delta}|y^\epsilon_s - \hat{y}^\epsilon_s|. \tag{A.42}$$

Hence, we obtain from (A.38) that

$$|K_1| \le C\delta e^{C\frac{\delta}{N\epsilon}} \tag{A.43}$$

Moreover, we obtain from Conditions 2.1.2 and 2.1.3 that

$$|K_2| \le CE\big(\frac{\delta}{N\epsilon}\big) \tag{A.44}$$

This concludes the proof of Lemma A.1.2. □

Lemma A.1.3. *We have for $m \in \mathbb{N}^*$*

$$\big|J_2(k)\big| \le C\delta k\Big(m\delta + E(\frac{m\tau}{\epsilon}) + \big(\frac{\tau}{\epsilon} + m\delta + m(\frac{\tau}{\epsilon})^2\big)e^{C\frac{m\tau}{\epsilon}}\Big). \tag{A.45}$$

Proof. Let $m \in \mathbb{N}^*$. Define $(\tilde{x}_s, \tilde{y}_s)$ such that for $j \in \mathbb{N}^*$, $n \in \mathbb{N}^*$,

$$\begin{cases} \frac{d\tilde{x}_s}{dt} = g(\tilde{x}_s,\tilde{y}_s) \quad \text{for} \quad jm\delta \le s < (j+1)m\delta \\ \frac{d\tilde{y}_s}{dt} = \frac{1}{\epsilon}f(\tilde{x}_s,\tilde{y}_s) \quad \text{for} \quad n\delta \le s < n\delta+\tau \\ \tilde{y}_s = \tilde{y}_{n\delta+\tau} \quad \text{for} \quad n\delta+\tau \le s < (n+1)\delta \\ \tilde{y}_{(n+1)\delta} = \tilde{y}_{n\delta+\tau} \quad \text{for} \quad n+1 \ne jm \\ (\tilde{x}_{jm},\tilde{y}_{jm}) = (\bar{x}_{jm\delta},\bar{y}_{jm\delta}) \end{cases} . \tag{A.46}$$

Define $\tilde{y}^a_s$ by

$$\begin{cases} \frac{d\tilde{y}^a_t}{dt} = \frac{1}{\epsilon} f(\bar{x}_{jm\delta}, \tilde{y}^a_t) \quad \text{for} \quad jm\tau \le t < (j+1)m\tau \\ \tilde{y}^a_{jm\tau} = \bar{y}_{jm\delta} \end{cases}, \tag{A.47}$$

and define $\tilde{x}^a_n$ by

$$\tilde{x}^a_n = \bar{x}_{jm\delta} \quad \text{for} \quad jm \le n < (j+1)m. \tag{A.48}$$

Observe that

$$J_2(k) = K_3 + K_4 + K_5 + K_6 + K_7 \tag{A.49}$$

with

$$K_3 := \sum_{n=0}^{k-1} \big(\int_{n\delta}^{(n+1)\delta} g(\tilde{x}_s, \tilde{y}_s)\, ds - \delta g(\bar{x}_{n\delta}, \bar{y}_{n\delta}) \big), \tag{A.50}$$

$$K_4 := \sum_{n=0}^{k-1} \delta \big(\frac{1}{\tau} \int_{n\tau}^{(n+1)\tau} g(\tilde{x}^a_n, \tilde{y}^a_s)\, ds - \frac{1}{\delta} \int_{n\delta}^{(n+1)\delta} g(\tilde{x}_s, \tilde{y}_s)\, ds \big), \tag{A.51}$$

$$K_5 := \frac{\delta}{\tau} \sum_{n=0}^{k-1} \big(\tau \bar{g}(\tilde{x}^a_n) - \int_{n\tau}^{(n+1)\tau} g(\tilde{x}^a_n, \tilde{y}^a_s)\, ds \big), \tag{A.52}$$

$$K_6 := \delta \sum_{n=0}^{k-1} \big(\bar{g}(\bar{x}_{n\delta}) - \bar{g}(\tilde{x}^a_n) \big). \tag{A.53}$$

Using the regularity of g we obtain

$$|K_6| \le \delta k C \delta m. \tag{A.54}$$

Arranging the right hand side of (A.51) into groups of m terms corresponding to the intervals of (A.47) we obtain, from Condition 2.1.2 and Condition 2.1.3, that

$$|K_5| \le Ck\delta E(\frac{m\tau}{\epsilon}). \tag{A.55}$$

Using (A.48) and the regularity of f and g we obtain the following inequality

$$|\tilde{y}^a_{\frac{\delta}{\tau}t} - \tilde{y}_t| \le Cm\delta e^{C\frac{m\tau}{\epsilon}}. \tag{A.56}$$

It follows that

$$|K_4| \leq C\delta km\delta e^{C\frac{m\tau}{\epsilon}}. \tag{A.57}$$

Similarly, using (A.7) and (A.8), we obtain the following inequalities

$$|\tilde{y}_{n\delta} - \bar{y}_{n\delta}| \leq C(\frac{\tau}{\epsilon} + m\delta + m(\frac{\tau}{\epsilon})^2)\frac{m\tau}{\epsilon}e^{C\frac{m\tau}{\epsilon}}, \tag{A.58}$$

$$|\tilde{x}_{n\delta} - \bar{x}_{n\delta}| \leq Cm\big(\delta + (\frac{\tau}{\epsilon})^2\big). \tag{A.59}$$

It follows that

$$|K_3| \leq C\delta k\big(\frac{\tau}{\epsilon} + m\delta + m(\frac{\tau}{\epsilon})^2\big)e^{C\frac{m\tau}{\epsilon}}. \tag{A.60}$$

This concludes the proof of Lemma A.1.3. □

Combining Lemma A.1.1, A.1.2 and A.1.3 we have obtained that

$$\begin{aligned} |x^{\epsilon}_{n\delta} - \bar{x}_{n\delta}| \leq & Ce^{C\delta n}\Big(\delta + \big(\frac{\tau}{\epsilon}\big)^2\frac{1}{\delta} + \delta e^{C\frac{\delta}{N\epsilon}} + E\big(\frac{\delta}{N\epsilon}\big) + E(\frac{m\tau}{\epsilon}) \\ & + \big(\frac{\tau}{\epsilon} + m\delta + m(\frac{\tau}{\epsilon})^2\big)e^{C\frac{m\tau}{\epsilon}}\Big) \end{aligned} \tag{A.61}$$

Choosing N such that $e^{C\frac{\delta}{N\epsilon}} \sim \delta^{-\frac{1}{2}}$ (observe that we need $\epsilon \leq \delta/(-C\ln\delta)$) and m such that $\frac{m\tau}{\epsilon}e^{C\frac{m\tau}{\epsilon}} \sim \big(\frac{\delta\epsilon}{\tau} + \frac{\tau}{\epsilon}\big)^{-\frac{1}{2}}$ we obtain for $\frac{\delta\epsilon}{\tau} + \frac{\tau}{\epsilon} \leq 1$ that

$$\begin{aligned} |x^{\epsilon}_{n\delta} - \bar{x}_{n\delta}| \leq & Ce^{C\delta n}\Bigg(\sqrt{\delta} + \big(\frac{\tau}{\epsilon}\big)^2\frac{1}{\delta} + E\big(\frac{1}{C}\ln\frac{1}{\delta}\big) \\ & + \big(\frac{\delta\epsilon}{\tau}\big)^{\frac{1}{2}} + \big(\frac{\tau}{\epsilon}\big)^{\frac{1}{2}} + E\Big(\frac{1}{C}\ln\Big(\big(\frac{\delta\epsilon}{\tau} + \frac{\tau}{\epsilon}\big)^{-1}\Big)\Big)\Bigg) \end{aligned} \tag{A.62}$$

This concludes the proof of inequality (A.2). The proof of (A.3) is similar and is also a consequence of (A.2).

A.2 Proof of Theorem 2.3.1

Define the process $t \mapsto (\bar{x}_t, \bar{y}_t)$ by

$$(\bar{x}_t, \bar{y}_t) := \eta(\bar{u}_t). \tag{A.63}$$

It follows from the regularity of η that it is sufficient to prove the F-convergence of $(\bar{x}_t, \bar{y}_t)$ towards $\delta_{X_t} \otimes \mu(X_t, dy)$. Now define ψ^ϵ_τ by

$$\psi^\epsilon_\tau(x, y, \omega) := \eta \circ \theta^\epsilon_\tau(., \omega) \circ \eta^{-1}(x, y), \tag{A.64}$$

Define ψ^g_h by

$$\psi^g_h(x, y, \omega) := \eta \circ \theta^G_h(., \omega) \circ \eta^{-1}(x, y). \tag{A.65}$$

Proposition A.2.1. *The vector fields f, g and matrix fields σ, Q associated with the system* (2.60) *are uniformly bounded and Lipschitz continuous. We also have*

$$\begin{cases} (\bar{x}_0, \bar{y}_0) = \eta(u_0) \\ (\bar{x}_{(k+1)\delta}, \bar{y}_{(k+1)\delta}) = \psi^g_{\delta-\tau}(., \omega'_k) \circ \psi^\epsilon_\tau\big((\bar{x}_{k\delta}, \bar{y}_{k\delta}), \omega_k\big) \\ (\bar{x}_t, \bar{y}_t) = (\bar{x}_{k\delta}, \bar{y}_{k\delta}) \quad \textit{for} \quad k\delta \le t < (k+1)\delta \end{cases} \tag{A.66}$$

where ω_k, ω'_k are i.i.d. samples from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Moreover, there exists $C > 0$ and and d-dimensional centered Gaussian vectors $\xi'(\omega)$, $\xi''(\omega)$ with identity covariance matrices such that for $h \le h_0$ and $\frac{\tau}{\epsilon} \le \tau_0$ we have

$$\left(\mathbb{E}\Big[\big|\psi^g_h(x, y, \omega) - (x, y) - h\big(g(x, y), 0\big) - \sqrt{h}\big(\sigma(x, y)\xi'(\omega), 0\big)\big|^2\Big]\right)^{\frac{1}{2}} \le Ch^{\frac{3}{2}}, \tag{A.67}$$

$$\left(\mathbb{E}\Big[\big|\psi^\epsilon_\tau(x, y, \omega) - (x, y) - \tau\big(g(x, y), 0\big) - \frac{\tau}{\epsilon}\big(0, f(x, y)\big) - \sqrt{\tau}\big(\sigma(x, y)\xi''(\omega), 0\big) - \sqrt{\frac{\tau}{\epsilon}}\big(0, Q(x, y)\xi''(\omega)\big)\big|^2\Big]\right)^{\frac{1}{2}} \le C\big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}}. \tag{A.68}$$

Proof. Since $(x,y)=\eta(u)$, we obtain from (2.59) and Itô's formula

$$\begin{aligned} dx =&\big((G+\frac{1}{\epsilon}F)\nabla\eta^x\circ\eta^{-1}(x,y)\big)\,dt+\big(\nabla\eta^x(H+\frac{1}{\sqrt{\epsilon}}K)\big)\circ\eta^{-1}(x,y)\,dW_t\\ &+\frac{1}{2}\sum_{ij}\partial_i\partial_j\eta^x\big((H+\frac{1}{\sqrt{\epsilon}}K)(H+\frac{1}{\sqrt{\epsilon}}K)^T\big)_{ij}\,dts \end{aligned} \tag{A.69}$$

$$\begin{aligned} dy =&\big((G+\frac{1}{\epsilon}F)\nabla\eta^y\circ\eta^{-1}(x,y)\big)\,dt+\big(\nabla\eta^y(H+\frac{1}{\sqrt{\epsilon}}K)\big)\circ\eta^{-1}(x,y)\,dW_t\\ &+\Big(\frac{1}{2}\sum_{ij}\partial_i\partial_j\eta^y\big((H+\frac{1}{\sqrt{\epsilon}}K)(H+\frac{1}{\sqrt{\epsilon}}K)^T\big)_{ij}\Big)\circ\eta^{-1}\,dt. \end{aligned} \tag{A.70}$$

Hence we deduce from (2.60) in Condition 2.3.1 that

$$g(x,y)=\Big(G\nabla\eta^x+\frac{1}{2}\sum_{ij}\partial_i\partial_j\eta^x(HH^T)_{ij}\Big)\circ\eta^{-1}(x,y) \tag{A.71}$$

$$\sigma(x,y)=\big(\nabla\eta^xH\big)\circ\eta^{-1}(x,y) \tag{A.72}$$

$$f(x,y)=\Big(F\nabla\eta^y+\frac{1}{2}\sum_{ij}\partial_i\partial_j\eta^y(KK^T)_{ij}\Big)\circ\eta^{-1}(x,y) \tag{A.73}$$

$$Q(x,y)=\big(\nabla\eta^yK\big)\circ\eta^{-1}(x,y). \tag{A.74}$$

Remark A.2.1. *Observe that* (2.60) *in Condition 2.3.1 requires that*

$$F\nabla\eta^x=0,\quad G\nabla\eta^y=0, \tag{A.75}$$

$$\sum_{ij}\partial_i\partial_j\eta^x\big(KK^T\big)_{ij}=0, \tag{A.76}$$

$$\sum_{ij}\partial_i\partial_j\eta^y\big(HH^T\big)_{ij}=0, \tag{A.77}$$

$$\sum_{ij}\partial_i\partial_j\eta^x\big(KH^T+HK^T\big)_{ij}=0, \tag{A.78}$$

and

$$\sum_{ij}\partial_i\partial_j\eta^y\big(KH^T+HK^T\big)_{ij}=0. \tag{A.79}$$

(A.78) *and* (A.79) *are satisfied if* KH^T *is skew-symmetric. One particular case*

could be, of course, $KH^T = 0$, which translates into the fact that for all u the ranges of $H(u)$ and $K(u)$ are orthogonal, i.e., the noise with amplitude $1/\sqrt{\epsilon}$ is applied to degrees of freedom orthogonal to those with $\mathcal{O}(1)$ noise.

We deduce the regularity of f, g, σ and Q from the regularity of G, F, H, K and η. (A.6) is a direct consequence of the definition of ψ_τ^ϵ and ψ_h^g and (A.66). Now observe that

$$\begin{aligned}
&\psi_\tau^\epsilon(x,y,\omega) - (x,y) - \tau\big(g(x,y),0\big) - \frac{\tau}{\epsilon}\big(0,f(x,y)\big) - \sqrt{\tau}\big(\sigma(x,y)\xi'(\omega),0\big)\\
&- \sqrt{\frac{\tau}{\epsilon}}\big(0,Q(x,y)\xi'(\omega)\big) = \Big(\eta\circ\theta_\tau^\epsilon - \eta - \tau\big(G\nabla\eta^x + \frac{1}{2}\sum_{ij}\partial_i\partial_j\eta^x(HH^T)_{ij},0\big)\\
&- \frac{\tau}{\epsilon}\big(0,F\nabla\eta^y + \frac{1}{2}\sum_{ij}\partial_i\partial_j\eta^y(KK^T)_{ij}\big) - \sqrt{\tau}\big(\nabla\eta^x H\xi'(\omega),0\big)\\
&- \sqrt{\frac{\tau}{\epsilon}}\big(0,\nabla\eta^y K\xi'(\omega)\big)\Big)\circ\eta^{-1}(x,y).
\end{aligned} \tag{A.80}$$

Using (A.75), (A.76), (A.77), (A.78) and (A.79), the Taylor-Ito expansion of $\eta\circ\theta_\tau^\epsilon$, the regularity of η, and setting ξ' equal to ξ defined in (2.72), we obtain (A.68). The proof of (A.67) is similar. □

It follows from Proposition A.2.1 that it is sufficient to prove Theorem 2.3.1 in the situation where η is the identity diffeomorphism. More precisely the F-convergence of $\bar{u}_t$ is a consequence of the F-convergence of $(\bar{x}_t,\bar{y}_t)$ and the regularity of η.

Let $x\mapsto\varphi(x)$ be a function with continuous and bounded derivatives up to order 3. Let us prove the following lemma.

Lemma A.2.1. *We have*

$$\begin{aligned}
&\mathbb{E}\big[\varphi(\bar{x}_{(n+1)\delta})\big] - \mathbb{E}\big[\varphi(\bar{x}_{n\delta})\big] =\\
&\qquad\qquad \delta\mathbb{E}\Big[g(\bar{x}_{n\delta},\bar{y}_{n\delta})\nabla\varphi(\bar{x}_{n\delta}) + \sigma\sigma^T(\bar{x}_{n\delta},\bar{y}_{n\delta}) : \operatorname{Hess}\varphi(\bar{x}_{n\delta})\Big] + I_0
\end{aligned} \tag{A.81}$$

with

$$|I_0| \leq C\Big(\delta^{\frac{3}{2}} + \big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}}\Big). \tag{A.82}$$

Proof. Write $(\bar{x}_{n\delta+\tau}, \bar{y}_{n\delta+\tau}) := \psi_\tau^\epsilon(\bar{x}_{n\delta}, \bar{y}_{n\delta}, \omega_n)$. Using (A.68), we obtain that there exists an $\mathcal{N}(0,1)$ random vector ξ_n, independent from $(\bar{x}_{n\delta}, \bar{y}_{n\delta})$, such that

$$\bar{x}_{n\delta+\tau} - \bar{x}_{n\delta} = g(\bar{x}_{n\delta})\tau + \sqrt{\tau}\sigma(\bar{x}_{n\delta}, \bar{y}_{n\delta})\xi_n + I_1 \tag{A.83}$$

with

$$\big(\mathbb{E}[(I_1)^2]\big)^{\frac{1}{2}} \leq C\big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}}. \tag{A.84}$$

Hence

$$\begin{aligned}\Big|\mathbb{E}\big[\varphi(\bar{x}_{n\delta+\tau})\big] - \mathbb{E}\big[\varphi(\bar{x}_{n\delta})\big] - \tau\mathbb{E}\Big[g(\bar{x}_{n\delta}, \bar{y}_{n\delta})\nabla\varphi(\bar{x}_{n\delta}) \\ + \sigma\sigma^T(\bar{x}_{n\delta}, \bar{y}_{n\delta}) : \mathrm{Hess}\,\varphi(\bar{x}_{n\delta})\Big]\Big| \leq C\big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}}\end{aligned} \tag{A.85}$$

Similarly, using (A.67), we obtain that there exists an $\mathcal{N}(0,1)$ random vector ξ_n', independent from $(\bar{x}_{n\delta+\tau}, \bar{y}_{n\delta+\tau})$, such that

$$\bar{x}_{(n+1)\delta} - \bar{x}_{n\delta+\tau} = g(\bar{x}_{n\delta+\tau}, \bar{y}_{n\delta+\tau})(\delta - \tau) + \sigma(\bar{x}_{n\delta+\tau}, \bar{y}_{n\delta+\tau})\sqrt{\delta-\tau}\xi_n' + I_2 \tag{A.86}$$

with

$$\big(\mathbb{E}[(I_2)^2]\big)^{\frac{1}{2}} \leq C(\delta - \tau)^{\frac{3}{2}}. \tag{A.87}$$

Whence

$$\begin{aligned}\Big|\mathbb{E}\big[\varphi(\bar{x}_{(n+1)\delta})\big] - \mathbb{E}\big[\varphi(\bar{x}_{n\delta+\tau})\big] - (\delta-\tau)\mathbb{E}\Big[g(\bar{x}_{n\delta+\tau},\bar{y}_{n\delta+\tau})\nabla\varphi(\bar{x}_{n\delta+\tau}) \\ + \sigma\sigma^T(\bar{x}_{n\delta+\tau},\bar{y}_{n\delta+\tau}) : \operatorname{Hess}\varphi(\bar{x}_{n\delta+\tau})\Big]\Big| \le C\big(\delta-\tau\big)^{\frac{3}{2}}.\end{aligned} \tag{A.88}$$

Using the regularity of σ, we obtain that

$$\Big(\mathbb{E}\Big[\big|\sigma(\bar{x}_{n\delta+\tau},\bar{y}_{(n+1)\delta}) - \sigma(\bar{x}_{n\delta},\bar{y}_{n\delta})\big|^2\Big]\Big)^{\frac{1}{2}} \le C\big(\delta^{\frac{1}{2}} + \sqrt{\frac{\tau}{\epsilon}}\big). \tag{A.89}$$

The proof of (A.81) follows from (A.68), (A.85), (A.88), (A.89) and the regularity of g and φ. □

Lemma A.2.2. *We have*

$$\Big|\frac{\mathbb{E}\big[\varphi(\bar{x}_{n\delta})\big] - \varphi(x_0)}{n\delta} - L\varphi(x_0)\Big| \le J_5 \tag{A.90}$$

with (for $\delta \le C\tau/\epsilon$)

$$|J_5| \le C\Big(\big(\frac{\delta\epsilon}{\tau}\big)^{\frac{1}{4}} + \big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}}\frac{1}{\delta} + \sqrt{\frac{\tau}{\epsilon}}\Big) + CE\big(\frac{1}{C}\ln\frac{\tau}{\delta\epsilon}\big). \tag{A.91}$$

Proof. Define $\hat{B}_t$ by $\hat{B}_0 = 0$ and

$$\hat{B}_t - \hat{B}_{n\tau} = B_{n\delta+t} - B_{n\delta} \quad \text{for} \quad n\tau \le t \le (n+1)\tau. \tag{A.92}$$

Define $\tilde{y}_s$ by $\tilde{y}_0 = y_0$ and

$$d\tilde{y}_t = \frac{1}{\epsilon}f(x_0,\tilde{y}_t)\,dt + \frac{1}{\sqrt{\epsilon}}Q(x_0,\tilde{y}_t)d\hat{B}_t. \tag{A.93}$$

Write

$$\bar{g}(x_0) := \int g(x_0,y)\,\mu(x_0,dy). \tag{A.94}$$

Using Lemma A.2.1 we obtain

$$\frac{\mathbb{E}\big[\varphi(\bar{x}_{n\delta})\big] - \varphi(x_0)}{n\delta} = L\varphi(x_0) + J_1 + J_2 + J_3 + J_4, \tag{A.95}$$

with

$$L\varphi(x_0) := \bar{g}(x_0)\nabla\varphi(x_0) + \bar{\sigma}\bar{\sigma}^T(x_0) : \mathrm{Hess}\,\varphi(x_0), \tag{A.96}$$

$$\begin{aligned} J_1 =& \frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}\Big[g(\bar{x}_{k\delta},\bar{y}_{k\delta})\nabla\varphi(\bar{x}_{k\delta}) + \sigma\sigma^T(\bar{x}_{k\delta},\bar{y}_{k\delta}) : \mathrm{Hess}\,\varphi(\bar{x}_{k\delta})\Big] \\ & - \frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}\Big[g(\bar{x}_0,\bar{y}_{k\delta})\nabla\varphi(\bar{x}_0) + \sigma\sigma^T(\bar{x}_0,\bar{y}_{k\delta}) : \mathrm{Hess}\,\varphi(\bar{x}_0)\Big], \end{aligned} \tag{A.97}$$

$$\begin{aligned} J_2 =& \frac{1}{n}\sum_{k=0}^{n-1}\Big(\mathbb{E}\Big[g(\bar{x}_0,\bar{y}_{k\delta})\nabla\varphi(\bar{x}_0) + \sigma\sigma^T(\bar{x}_0,\bar{y}_{k\delta}) : \mathrm{Hess}\,\varphi(\bar{x}_0)\Big] \\ & - \frac{1}{\tau}\int_{k\tau}^{(k+1)\tau}\mathbb{E}\Big[g(x_0,\tilde{y}_s)\nabla\varphi(x_0) + \sigma\sigma^T(x_0,\tilde{y}_s) : \mathrm{Hess}\,\varphi(x_0)\Big]\,ds\Big), \end{aligned} \tag{A.98}$$

$$J_3 = \frac{1}{n\tau}\int_0^{n\tau}\mathbb{E}\Big[g(x_0,\tilde{y}_s)\nabla\varphi(x_0) + \sigma\sigma^T(x_0,\tilde{y}_s) : \mathrm{Hess}\,\varphi(x_0)\Big]\,ds - L\varphi(x_0), \tag{A.99}$$

$$|J_4| \le C\Big(\delta^{\frac{1}{2}} + \big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}}\frac{1}{\delta}\Big). \tag{A.100}$$

Using the regularity of σ, g, φ, (A.6) and (A.7) we obtain

$$|J_1| \le C\Big((n\delta)^{\frac{1}{2}} + n\delta + n\big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}}\Big). \tag{A.101}$$

Using Property 3 of Condition 2.3.1 and Property 3 of Condition 2.3.2 we obtain

$$|J_3| \le CE(\frac{n\tau}{\epsilon}). \tag{A.102}$$

Using (A.67) and (A.68), we obtain the following inequality

$$\Big(\mathbb{E}\Big[\big|\bar{y}_{n\delta}-\tilde{y}_{n\tau}\big|^2\Big]\Big)^{\frac{1}{2}} \le C\Big(\sqrt{\frac{\tau}{\epsilon}}+(n\delta)^{\frac{1}{2}}+n\delta+n\big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}}\Big)\frac{n\tau}{\epsilon}e^{C\frac{n\tau}{\epsilon}}, \tag{A.103}$$

which leads to

$$|J_2| \le C\Big(\sqrt{\frac{\tau}{\epsilon}}+(n\delta)^{\frac{1}{2}}+n\delta+n\big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}}\Big)e^{C\frac{n\tau}{\epsilon}}. \tag{A.104}$$

Hence, we have obtained

$$\Big|\frac{\mathbb{E}\big[\varphi(\bar{x}_{n\delta})\big]-\varphi(x_0)}{n\delta}-L\varphi(x_0)\Big| \le J_5, \tag{A.105}$$

with

$$|J_5| \le C\Big(\sqrt{\frac{\tau}{\epsilon}}+(n\delta)^{\frac{1}{2}}+n\delta+n\big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}}\Big)e^{C\frac{n\tau}{\epsilon}}+E(\frac{n\tau}{\epsilon})+C\big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}}\frac{1}{\delta}. \tag{A.106}$$

Choosing n such that $\sqrt{\frac{n\tau}{\epsilon}}e^{C\frac{n\tau}{\epsilon}} \sim \big(\frac{\tau}{\epsilon\delta}\big)^{\frac{1}{4}}$ we obtain (A.91) for $\delta \le C\tau/\epsilon$. □

We now combine Lemma A.2.2 with Theorem 1 of Chapter 2 of [258] which states that the uniform convergence (in x_0, y_0) of $\frac{\mathbb{E}[\varphi(\bar{x}_{n\delta})]-\varphi(x_0)}{n\delta}$ to $L\varphi(x_0)$ as $\epsilon \downarrow 0$, $\tau \le \delta$, $\frac{\tau}{\epsilon} \downarrow 0$, $\frac{\delta\epsilon}{\tau} \downarrow 0$ and $\big(\frac{\tau}{\epsilon}\big)^{\frac{3}{2}}\frac{1}{\delta} \downarrow 0$ implies the convergence in distribution of $\bar{x}_{n\delta}$ to the Markov process generated by L.

The F-convergence of $(\bar{x}_t, \bar{y}_t)$ can be deduced from the convergence in distribution of $\bar{x}_t$ and (2.62) of Condition 2.3.1. The proof follows the same lines as above, which will not be repeated here.

A.3 Proof of Theorem 4.2.1

Throughout this section Condition 4.2.1 is assumed. For concise writing, we also abuse the notation $O(x^n)$, which indicates some entity whose norm $\le Cx^n$, where C is a constant that does not depend on ϵ^{-1}.

Since ϵ is very small, the following inequalities for converting between scaled energy norm and two-norm can be easily obtained:

Proposition A.3.1. *Let* $x = \begin{bmatrix} q \\ p \end{bmatrix}$ *be any vector, then*

$$\epsilon^{1/2}\|\sqrt{K}\|_2^{-1}\|x\|_2 = \|\Omega\|_2^{-1}\|x\|_2 \leq \|x\|_E \leq \|x\|_2 \tag{A.107}$$

$$\| \begin{bmatrix} 0 \\ p \end{bmatrix} \|_E \leq \|\Omega^{-1}\|_2\| \begin{bmatrix} 0 \\ p \end{bmatrix} \|_2 = \epsilon^{1/2}\|\sqrt{K}^{-1}\|_2\|x\|_2 \tag{A.108}$$

Also, vector-norm-induced matrix norms satisfy

$$\| \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \|_E \triangleq \sup \frac{\|Mx\|_E}{\|x\|_E} = \| \begin{bmatrix} M_{11} & M_{12}\Omega \\ \Omega^{-1}M_{21} & \Omega^{-1}M_{22}\Omega \end{bmatrix} \|_2 \tag{A.109}$$

Lemma A.3.1. *Let* $B(s) = \begin{bmatrix} B_{11}(s) & B_{12}(s) \\ B_{21}(s) & B_{22}(s) \end{bmatrix} = \exp(s \begin{bmatrix} 0 & I \\ -\epsilon^{-1}K & c \end{bmatrix})$, *and* $Rq(s)$ *be the* $Rq_k(H)$ *defined in Integrator 4.2.2 with* $H = s$ *and arbitrary* k, *then*

$$\|B_{11}(s)\|_2 \leq 1 \tag{A.110}$$

$$\|B_{22}(s)\|_2 \leq 1 \tag{A.111}$$

$$\|B_{12}(s)\|_2 \leq |s| \tag{A.112}$$

$$\epsilon\|B_{21}(s)\|_2 \leq C_K|s| \tag{A.113}$$

$$\epsilon^{1/2}\|B_{11}(s) - I\|_2 \leq C_K|s| \tag{A.114}$$

$$\epsilon^{1/2}\|B_{22}(s) - I\|_2 \leq C_c|s| \tag{A.115}$$

$$\mathbb{E}\|Rq(s)\|_2^2 \leq \frac{1}{3}\|\sigma\|_2^2|s|^3 \tag{A.116}$$

$$\epsilon^{1/2}\|B(s) - I\|_2 \leq C_{Kc}|s| \tag{A.117}$$

$$\epsilon^{1/2}\|B(s) - I\|_E \leq C_{Kc}|s| \tag{A.118}$$

where C_K, C_c *and* C_{Kc} *are some positive real constants (may indicate different values in different inequalities), respectively dependent on* K, $\sqrt{\epsilon}c$, K *and* $\sqrt{\epsilon}c$ *but independent of* ϵ^{-1}.

Proof. Since c and K commute, they can be diagonalized simultaneously (see for instance [145]). By the theory of linear ordinary differential equations (see for instance [228]), one can hence diagonalize B_{11}, B_{12}, B_{21}, B_{22} simultaneously. Since each diagonal element can be investigated individually, assume without loss of generality that $\Omega = [\omega]_{ij} = \epsilon^{-1/2}\sqrt{K}$ and c are both scalars, and use the notation of scalar ω and scalar c thereafter.

Denote the damping ratio by $\zeta = \frac{c}{\omega}$. The solution to damped harmonic oscillator can be analytically obtained, and hence components of the flow operator B_{11},B_{12},B_{21},B_{22} as well.

When $\zeta < 1$ i.e., underdamping, which is usually the case since ω is large

$$B_{11}(s) = e^{-\omega\zeta s}(cos(\omega\sqrt{1-\zeta^2}s) + \frac{\zeta}{\sqrt{1-\zeta^2}}sin(\sqrt{1-\zeta^2}s)) \quad (A.119)$$

$$B_{12}(s) = \frac{e^{-\omega\zeta s}sin(\omega\sqrt{1-\zeta^2}s)}{\omega\sqrt{1-\zeta^2}} \quad (A.120)$$

$$B_{21}(s) = -\omega\frac{e^{-\omega\zeta s}sin(\omega\sqrt{1-\zeta^2}s)}{\sqrt{1-\zeta^2}} \quad (A.121)$$

$$B_{22}(s) = e^{-\omega\zeta s}(cos(\omega\sqrt{1-\zeta^2}s) - \frac{\zeta}{\sqrt{1-\zeta^2}}sin(\sqrt{1-\zeta^2}s)) \quad (A.122)$$

When $\zeta = 1$ i.e., critical damping,

$$B_{11}(s) = e^{-\omega s}(1+\omega s) \quad (A.123)$$

$$B_{12}(s) = e^{-\omega s}s \quad (A.124)$$

$$B_{21}(s) = -\omega^2 e^{-\omega t}t \quad (A.125)$$

$$B_{22}(s) = e^{-\omega s}(1-\omega s) \quad (A.126)$$

When $\zeta > 1$ i.e., over damping,

$$A(s) \triangleq e^{\omega s(-\zeta-\sqrt{\zeta^2-1})} \tag{A.127}$$

$$B(s) \triangleq e^{\omega s(-\zeta+\sqrt{\zeta^2-1})} \tag{A.128}$$

$$B_{11}(s) = \frac{\zeta(B-A)+\sqrt{\zeta^2-1}(A+B)}{2\sqrt{\zeta^2-1}} \tag{A.129}$$

$$B_{12}(s) = \frac{-A+B}{2\omega\sqrt{\zeta^2-1}} \tag{A.130}$$

$$B_{21}(s) = \frac{\omega(A-B)}{2\sqrt{\zeta^2-1}} \tag{A.131}$$

$$B_{22}(s) = \frac{\zeta(A-B)+\sqrt{\zeta^2-1}(A+B)}{2\sqrt{\zeta^2-1}} \tag{A.132}$$

$$\tag{A.133}$$

By routine investigations on local extremes using calculus, it can be shown in all three cases that

$$\|B_{11}(s)\|_2 \leq 1 \tag{A.134}$$

$$\|B_{22}(s)\|_2 \leq 1 \tag{A.135}$$

$$\|B_{12}(s)\|_2 \leq s \tag{A.136}$$

$$\|B_{21}(s)\|_2 \leq \omega^2 s \tag{A.137}$$

$$\|B_{11}(s)-I\|_2 \leq \omega s \tag{A.138}$$

$$\|B_{22}(s)-I\|_2 \leq \begin{cases} \omega s & \zeta \leq 1 \\ 2\zeta\omega s & \zeta > 1 \end{cases} \tag{A.139}$$

When $\zeta > 1$, since $c = O(\epsilon^{-1/2})$ (Condition 4.2.1), $2\zeta\omega s = O(\epsilon^{-1/2})s$. Therefore $\epsilon^{1/2}\|B_{22}(s)-I\|_2 \leq C_c|s|$ always holds.

Also,

$$\begin{aligned}\mathbb{E}\|Rq(s)\|_2^2 &= \mathbb{E}\| \int_0^s B_{12}(t)\sigma dW_t\|_2^2 \\ &= \int_0^s \|\sigma B_{12}(t)\|_2^2 dt \leq \frac{1}{3}\|\sigma\|_2^2|s|^3 \end{aligned} \tag{A.140}$$

For a proof on norm bounds of the entire matrice we use only bounds of dimensionless block elements:

$$\|B - I\|_2 = \| \begin{bmatrix} B_{11} - I & B_{12} \\ B_{21} & B_{22} - I \end{bmatrix} \|_2 \tag{A.141}$$

$$\leq \| \begin{bmatrix} \Omega & 0 \\ 0 & \Omega \end{bmatrix} \|_2 \| \begin{bmatrix} \Omega^{-1}(B_{11} - I) & \Omega^{-1}B_{12} \\ \Omega^{-1}B_{21} & \Omega^{-1}(B_{22} - I) \end{bmatrix} \|_2 \tag{A.142}$$

$$= \epsilon^{-1/2} \| \begin{bmatrix} O(s) & \epsilon^{1/2}O(s) \\ \epsilon^{-1/2}O(s) & O(s) \end{bmatrix} \|_2 \tag{A.143}$$

It's easy to prove that for any scalar a

$$\| \begin{bmatrix} M_{11} & aM_{12} \\ M_{21} & M_{22} \end{bmatrix} \|_2 = \| \begin{bmatrix} M_{11} & M_{12} \\ aM_{21} & M_{22} \end{bmatrix} \|_2 \tag{A.144}$$

Therefore

$$\|B - I\|_2 = \epsilon^{-1/2} \| \begin{bmatrix} O(s) & O(s) \\ O(s) & O(s) \end{bmatrix} \|_2 = \epsilon^{-1/2}O(s) \tag{A.145}$$

Similarly,

$$\begin{aligned}
\|B-I\|_E &= \epsilon^{-1/2}\| \begin{bmatrix} B_{11}-I & B_{12}\Omega \\ \Omega^{-1}B_{21} & \Omega^{-1}B_{22}\Omega - I \end{bmatrix} \|_2 \\
&\leq \epsilon^{-1/2}\| \begin{bmatrix} O(s) & \epsilon^{1/2}\epsilon^{-1/2}O(s) \\ \epsilon^{-1/2}\epsilon^{1/2}O(s) & \epsilon^{1/2}O(s)\epsilon^{-1/2} \end{bmatrix} \|_2 \\
&= \epsilon^{-1/2}\| \begin{bmatrix} O(s) & O(s) \\ O(s) & O(s) \end{bmatrix} \|_2 \\
&= \epsilon^{-1/2}O(s) \qquad \text{(A.146)}
\end{aligned}$$

□

Remark A.3.1. *In the special case of $c = 0$, bounds of block elements can be easily obtained since*

$$\begin{aligned}
|cos(\omega s)| &\leq 1 \\
\epsilon K^{-1}|-\omega sin(\omega s)| &= |\frac{-\omega sin(\omega s)}{\omega^2}| \leq |s| \\
\epsilon^{1/2}\sqrt{K}^{-1}|cos(\omega s)-1| &= |-2sin^2(\omega s/2)/\omega| \leq |-2sin(\omega s/2)/\omega| \leq |s|
\end{aligned}$$

Lemma A.3.2. *The solution to the SDE*

$$dX = AXdt + f(X)dt + \Sigma dW_t$$

can be written in the following integral form:

$$X(t) = e^{At}X(0) + \int_0^t e^{A(t-s)}f(X(s))ds + \int_0^t e^{A(t-s)}\Sigma dW_s \qquad \text{(A.147)}$$

Proof. Let $Y(t) = e^{-At}X(t)$, then by Ito's formula and $dX = AXdt + f(X)dt + \Sigma dW_t$

$$dY = e^{-At}f(X(t))dt + e^{-At}\Sigma dW_t \qquad \text{(A.148)}$$

This in the integral form is

$$Y(t) = Y(0) + \int_0^t e^{-As} f(X(s))ds + \int_0^t e^{-As}\Sigma dW_s \tag{A.149}$$

Hence

$$X(t) = e^{At}X(0) + \int_0^t e^{A(t-s)} f(X(s))ds + \int_0^t e^{A(t-s)}\Sigma dW_s \tag{A.150}$$

□

Lemma A.3.3. *Consider two continuous stochastic dynamical systems, the original dynamics and the bridge dynamics:*

$$\begin{cases} dq &= pdt \\ dp &= -\epsilon^{-1}Kqdt - \nabla V(q)dt - cpdt + \sigma dW_t \\ q(0) &= q_0 \\ p(0) &= p_0 \end{cases} \tag{A.151}$$

$$\begin{cases} d\tilde{q} &= \tilde{p}dt \\ d\tilde{p} &= -\epsilon^{-1}K\tilde{q}dt - \nabla V(q_0)dt - c\tilde{p}dt + \sigma dW_t \\ \tilde{q}(0) &= q_0 \\ \tilde{p}(0) &= p_0 \end{cases} \tag{A.152}$$

Then $(\mathbb{E}\| \begin{bmatrix} \tilde{q}(H) \\ \tilde{p}(H) \end{bmatrix} - \begin{bmatrix} q(h) \\ p(h) \end{bmatrix} \|_E^2)^{1/2} \leq C|H|^{3/2}$*, where* C *is a positive constant independent of* ϵ^{-1} *but dependent on the scaleless elasticity matrix* K*, scaled damping coefficient* $\sqrt{\epsilon}c$*, amplitude of noise* σ*, and slow potential* $V(\cdot)$*.*

Proof. Rewrite the original dynamics (A.151) as

$$\begin{cases} dq &= pdt \\ dp &= \epsilon^{-1}Kqdt - \nabla V(q_0)dt + (\nabla V(q_0) - \nabla V(q))dt - cpdt + \sigma dW_t \\ q(0) &= q_0 \\ p(0) &= p_0 \end{cases} \tag{A.153}$$

Let $x(t) = \begin{bmatrix} q(t) \\ p(t) \end{bmatrix}$, $\tilde{x}(t) = \begin{bmatrix} \tilde{q}(t) \\ \tilde{p}(t) \end{bmatrix}$, $B(t) = \exp(t \begin{bmatrix} 0 & I \\ -\epsilon^{-1}K & -c \end{bmatrix})$, $b = \begin{bmatrix} 0 \\ -\nabla V(q_0) \end{bmatrix}$, $g(q,p) = g(x) = \begin{bmatrix} 0 \\ \nabla V(q_0) - \nabla V(q) \end{bmatrix}$, and $\Sigma = \begin{bmatrix} 0 \\ \sigma \end{bmatrix}$. Then by Lemma A.3.2 solutions to the original dynamics and bridge dynamics can be respectively written as:

$$\begin{aligned} x(t) &= B(t)x(0) + \int_0^t B(t-s)bds + \int_0^t B(t-s)\Sigma dW_s + \int_0^t B(t-s)g(x(s))ds \\ \tilde{x}(t) &= B(t)x(0) + \int_0^t B(t-s)bds + \int_0^t B(t-s)\Sigma dW_s \end{aligned} \tag{A.154}$$

Notice for any vector y and positive t that $\|B(t)y\|_E \leq \|y\|_E$, because energy is decaying in the system $\ddot{q} + c\dot{q} + \epsilon^{-1}Kq = 0$. Together with Cauchy-Schwarz we have

$$\begin{aligned} \mathbb{E}\|\tilde{x}(t) - x(t)\|_E^2 &= \mathbb{E}\| \int_0^t B(t-s)g(x(s))ds\|_E^2 \\ &\leq t\int_0^t \mathbb{E}\|B(t-s)g(x(s))\|_E^2 ds \\ &\leq t\int_0^t \mathbb{E}\|g(x(s))\|_E^2 ds \end{aligned} \tag{A.155}$$

By Condition 4.2.1, assume $\nabla V(\cdot)$ is Lipschitz continuous with coefficient L,

then almost surely

$$\begin{aligned}\|g(x(s))\|_E &= \| \begin{bmatrix} 0 \\ \nabla V(q_0) - \nabla V(q(s)) \end{bmatrix} \|_E \\ &\leq \sqrt{\epsilon}\|\sqrt{K}^{-1}\|_2 \|\nabla V(q_0) - \nabla V(q(s))\|_2 \\ &\leq L\sqrt{\epsilon}\|\sqrt{K}^{-1}\|_2 \|q(s) - q_0\|_2 \end{aligned} \tag{A.156}$$

Similarly, since

$$x(t) = B(t)x(0) + \int_0^t B(t-s) \begin{bmatrix} 0 \\ -\nabla V(q(s)) \end{bmatrix} ds + \int_0^t B(t-s)\Sigma dW_s \tag{A.157}$$

we have

$$\|x(s) - B(s)x_0 - \int_0^s B(s-t)\Sigma dW_t\|_2 \leq \int_0^s \|\nabla V(q(t))\|_2 dt \tag{A.158}$$

By Condition 4.2.1, $\nabla V(\cdot)$ is bounded, and hence the above is $O(s)$.

We now can bound (A.156) and therefore (A.155) with the aid of (A.158) and Lemma A.3.1:

$$\begin{aligned} & \mathbb{E}\|q(s) - q_0\|_2^2 \\ \leq\ & \mathbb{E}\|x(s) - x_0\|_2^2 \\ \leq\ & \mathbb{E}\left(\|x(s) - B(s)x_0 - \int_0^s B(t-s)\Sigma dW_s\|_2 + \|B(s)x_0 - x_0\|_2 + \| \int_0^s B(t-s)\Sigma dW_s\|_2 \right)^2 \\ \leq\ & 3\mathbb{E}\left(\|x(s) - B(s)x_0 - \int_0^s B(t-s)\Sigma dW_s\|_2^2 + \|B(s)x_0 - x_0\|_2^2 + \| \int_0^s B(t-s)\Sigma dW_s\|_2^2 \right) \\ =\ & 3\left(O(s^2) + \epsilon^{-1}O(s^2)\mathbb{E}\|x_0\|_2^2 + \int_0^s \sigma^2(B_{12}(t-s)^2 + B_{22}(t-s)^2)dt \right) \\ =\ & O(s^2) + \epsilon^{-1}O(s^2)\mathbb{E}\|x_0\|_2^2 + O(s^3) + O(s) \end{aligned} \tag{A.159}$$

By Condition 4.2.1, $\mathbb{E}\|x_0\|_2^2 = O(1)$. Therefore, the above expression is $\epsilon^{-1}O(s^2) + O(s)$.

This gives $\mathbb{E}\|g(\tilde{x}(s))\|_E^2 = O(s)$ independent of ϵ^{-1}, and eventually $\mathbb{E}\|\tilde{x}(h) -$

$x(h)\|_E^2 = O(h^3)$. □

Lemma A.3.4. *Consider the discrete stochastic dynamical system given by 1st-order SIM (Integrator 4.2.2):*

$$\begin{cases} q_H &= B_{11}(H)q_0 + B_{12}(H)p_0 + Rq(H) \\ p_H &= B_{21}(H)q_0 + B_{22}(H)p_0 + Rp(H) - H\nabla V(B_{11}(H)q_0 + B_{12}(H)p_0 + Rq(H)) \end{cases} \tag{A.160}$$

Then a comparison with bridge dynamics (A.152) *gives* $\mathbb{E}\|q_H - \tilde{q}(H)\|_2^2 \leq CH^4$ *and* $\mathbb{E}\|\Omega^{-1}(p_H - \tilde{p}(H))\|_2^2 \leq CH^4$*, and therefore*

$$(\mathbb{E}\| \begin{bmatrix} q_H \\ p_H \end{bmatrix} - \begin{bmatrix} \tilde{q}(H) \\ \tilde{p}(H) \end{bmatrix} \|_E^2)^{1/2} \leq CH^2 \tag{A.161}$$

where C's are positive constants independent of ϵ^{-1} *but dependent on scaleless elasticity matrix* K*, scaled damping coefficient* $\sqrt{\epsilon}c$*, amplitude of noise* σ*, and slow potential* $V(\cdot)$*.*

Proof. The exact solution to the bridge dynamics is

$$\begin{cases} \tilde{q}(H) &= B_{11}(H)q_0 + B_{12}(H)p_0 + \int_0^H B_{12}(s)(-\nabla V(q_0))ds + Rq(H) \\ \tilde{p}(H) &= B_{21}(H)q_0 + B_{22}(H)p_0 + \int_0^H B_{22}(s)(-\nabla V(q_0))ds + Rp(H) \end{cases} \tag{A.162}$$

Hence almost surely $\tilde{q}(H) - q_H = \int_0^H B_{12}(s)(-\nabla V(q_0))ds$.

Since $B_{12}(s) = O(s)$ by Lemma A.3.1, and $\mathbb{E}\| - \nabla V(q_0)\|_2^2$ is bounded by Condition (4.2.1), one gets

$$\begin{aligned} \mathbb{E}\|\tilde{q}(H) - q_H\|_2^2 &\leq H\int_0^H \mathbb{E}\|B_{12}(s)(-\nabla V(q_0))\|_2^2 ds \\ &\leq H\int_0^H \mathbb{E}(\|B_{12}(s)\|_2\| - \nabla V(q_0)\|_2)^2 ds \\ &= H\int_0^H O(s^2)\mathbb{E}\| - \nabla V(q_0)\|_2^2 ds \\ &= O(H^4) \end{aligned} \tag{A.163}$$

Investigation on p by applying Lemma A.3.1 and Condition 4.2.1 gives:

$$\begin{aligned}
& \mathbb{E}\|\Omega^{-1}(\tilde{p}(H) - p_H)\|_2^2 \\
= \; & \mathbb{E}\|\Omega^{-1}(\int_0^H B_{22}(s)ds(-\nabla V(q_0)) + H\nabla V(B_{11}(H)q_0 + B_{12}(H)p_0 + Rq(H)))\|_2^2 \\
= \; & \mathbb{E}\|\int_0^H \Omega^{-1}(B_{22}(s) - I)ds(-\nabla V(q_0)) + H\Omega^{-1}(\nabla V(B_{11}(H)q_0 + B_{12}(H)p_0 \\
& + Rq(H)) - \nabla V(q_0))\|_2^2 \\
\leq \; & 2\mathbb{E}[\|\int_0^H \Omega^{-1}(B_{22}(s) - I)ds(-\nabla V(q_0))\|_2^2 + \|H\Omega^{-1}(\nabla V(B_{11}(H)q_0 + B_{12}(H)p_0 \\
& + Rq(H)) - \nabla V(q_0))\|_2^2] \\
\leq \; & 2[H\int_0^H \mathbb{E}\|\Omega^{-1}(B_{22}(s) - I)(-\nabla V(q_0))\|_2^2 ds + \mathbb{E}\|H\Omega^{-1}(\nabla V(B_{11}(H)q_0 \\
& + B_{12}(H)p_0 + Rq(H)) - \nabla V(q_0))\|_2^2] \\
\leq \; & 2H[\int_0^H \|\Omega^{-1}(B_{22}(s) - I)\|_2^2 ds\mathbb{E}\|(-\nabla V(q_0))\|_2^2 + H\mathbb{E}\|\Omega^{-1}(\nabla V(B_{11}(H)q_0 \\
& + B_{12}(H)p_0 + Rq(H)) - \nabla V(q_0))\|_2] \\
\leq \; & 2H[O(H^3) + L^2H\mathbb{E}\|\Omega^{-1}(B_{11}(H)q_0 + B_{12}(H)p_0 + Rq(H) - q_0)\|_2^2] \\
\leq \; & 2H[O(H^3) + 3L^2H(\mathbb{E}\|\Omega^{-1}(B_{11}(H) - I)q_0\|_2^2 + \|\Omega^{-1}B_{12}(H)p_0\|_2^2 \\
& + \mathbb{E}\|\Omega^{-1}Rq(H)\|_2^2)] \\
\leq \; & 2H[O(H^3) + 3L^2H(\|\Omega^{-1}(B_{11}(H) - I)\|_2^2\mathbb{E}\|q_0\|_2^2 + \|B_{12}(H)\|_2^2\mathbb{E}\|p_0\|_2^2 \\
& + \mathbb{E}\|Rq(H)\|_2^2)] \\
\leq \; & 2H[O(H^3) + 3L^2H(O(H)^2\mathbb{E}\|q_0\|_2^2 + O(H)^2\mathbb{E}\|p_0\|_2^2 + O(H^3))] \\
= \; & O(H^4) \qquad\qquad (A.164)
\end{aligned}$$

Therefore $\mathbb{E}\| \begin{bmatrix} q_H \\ p_H \end{bmatrix} - \begin{bmatrix} \tilde{q}(H) \\ \tilde{p}(H) \end{bmatrix} \|_E^2 = O(H^4)$ independent of ϵ^{-1}. □

Lemma A.3.5. *Consider evolutions of different local initial conditions under the bridge dynamics:*

$$\begin{cases} d\tilde{q}_1 &= \tilde{p}_1 dt \\ d\tilde{p}_1 &= -\epsilon^{-1}K\tilde{q}_1 dt - \nabla V(\tilde{q}_1(0))dt - c\tilde{p}_1 dt + \sigma dW_t \end{cases} \qquad (A.165)$$

$$\begin{cases} d\tilde{q}_2 &= \tilde{p}_2 dt \\ d\tilde{p}_2 &= -\epsilon^{-1} K \tilde{q}_2 dt - \nabla V(\tilde{q}_2(0)) dt - c\tilde{p}_2 dt + \sigma dW_t \end{cases} \tag{A.166}$$

Denote by L the Lipschitz coefficient of $\nabla V(\cdot)$ (i.e., $\|\nabla V(a) - \nabla V(b)\|_2 \leq L\|a - b\|_2$), then almost surely

$$\| \begin{bmatrix} \tilde{q}_1(H) - \tilde{q}_2(H) \\ \tilde{p}_1(H) - \tilde{p}_2(H) \end{bmatrix} \|_E \leq (1 + HL) \| \begin{bmatrix} \tilde{q}_1(0) - \tilde{q}_2(0) \\ \tilde{p}_1(0) - \tilde{p}_2(0) \end{bmatrix} \|_E \tag{A.167}$$

Proof. Write out the solution to the bridge dynamics in integral form:

$$\begin{aligned} \begin{bmatrix} \tilde{q}_1(H) \\ \tilde{p}_1(H) \end{bmatrix} &= B(H) \begin{bmatrix} \tilde{q}_1(0) \\ \tilde{p}_1(0) \end{bmatrix} + \int_0^H B(H-s) \begin{bmatrix} 0 \\ -\nabla V(\tilde{q}_1(0)) \end{bmatrix} ds + \int_0^H B(H-s)\Sigma dW_s \\ \begin{bmatrix} \tilde{q}_2(H) \\ \tilde{p}_2(H) \end{bmatrix} &= B(H) \begin{bmatrix} \tilde{q}_2(0) \\ \tilde{p}_2(0) \end{bmatrix} + \int_0^H B(H-s) \begin{bmatrix} 0 \\ -\nabla V(\tilde{q}_2(0)) \end{bmatrix} ds + \int_0^H B(H-s)\Sigma dW_s \end{aligned} \tag{A.168}$$

Hence almost surely

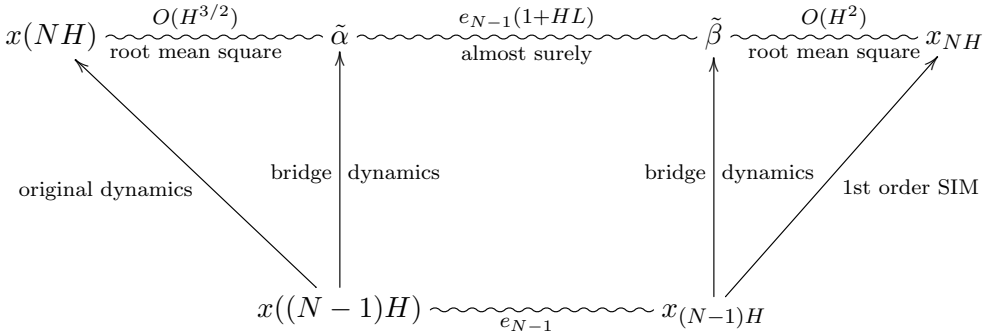
$$
\begin{aligned}
& \| \begin{bmatrix} \tilde{q}_1(H) - \tilde{q}_2(H) \\ \tilde{p}_1(H) - \tilde{p}_2(H) \end{bmatrix} \|_E \\
\leq \quad & \|B(H) \begin{bmatrix} \tilde{q}_1(0) - \tilde{q}_2(0) \\ \tilde{p}_1(0) - \tilde{p}_2(0) \end{bmatrix} \|_E + \int_0^H \|B(H-s) \begin{bmatrix} 0 \\ \nabla V(\tilde{q}_2(0)) - \nabla V(\tilde{q}_1(0)) \end{bmatrix} \|_E ds \\
\leq \quad & \| \begin{bmatrix} \tilde{q}_1(0) - \tilde{q}_2(0) \\ \tilde{p}_1(0) - \tilde{p}_2(0) \end{bmatrix} \|_E + H \| \begin{bmatrix} 0 \\ \nabla V(\tilde{q}_2(0)) - \nabla V(\tilde{q}_1(0)) \end{bmatrix} \|_E \\
\leq \quad & \| \begin{bmatrix} \tilde{q}_1(0) - \tilde{q}_2(0) \\ \tilde{p}_1(0) - \tilde{p}_2(0) \end{bmatrix} \|_E + HL \| \begin{bmatrix} 0 \\ \tilde{q}_2(0) - \tilde{q}_1(0) \end{bmatrix} \|_E \\
\leq \quad & \| \begin{bmatrix} \tilde{q}_1(0) - \tilde{q}_2(0) \\ \tilde{p}_1(0) - \tilde{p}_2(0) \end{bmatrix} \|_E + HL \| \begin{bmatrix} \tilde{q}_2(0) - \tilde{q}_1(0) \\ 0 \end{bmatrix} \|_E \\
\leq \quad & (1+HL) \| \begin{bmatrix} \tilde{q}_1(0) - \tilde{q}_2(0) \\ \tilde{p}_1(0) - \tilde{p}_2(0) \end{bmatrix} \|_E
\end{aligned} \tag{A.169}
$$

□

Remark A.3.2. *If the traditional method of investigating the Lipschitz coefficient of the vector field is employed to evolve the separation of local initial conditions, ϵ^{-1} will exhibit in the bound of separation. Instead we only looked at the soft part of the vector field and thence obtained a uniform bound.*

Theorem 4.2.1 (global error bound in energy norm).

Proof. **Atlas of error propagation:**



Let $e_N = (\mathbb{E}\|x(NH) - x_{NH}\|_E^2)^{1/2}$. Let $\tilde{\alpha}$ and $\tilde{\beta}$ be respectively the evolution of the real solution $x((N-1)H)$ and the numerical solution $x_{(N-1)H}$ by time H under the bridge dynamics (A.152).

Then by Lemma A.3.3 and A.3.4, there exist constants C_1 and C_2 independent of ϵ^{-1} such that

$$\begin{aligned} (\mathbb{E}\|X(NH) - \tilde{\alpha}\|_E^2)^{1/2} &\leq C_1 H^{3/2} \\ (\mathbb{E}\|\tilde{\beta} - X_{Nh}\|_E^2)^{1/2} &\leq C_2 H^2 \end{aligned} \tag{A.170}$$

Also since $\|\tilde{\alpha}-\tilde{\beta}\|_E \leq (1+HL)\|x((N-1)h)-x_{(N-1)h}\|_E$ almost surely (Lemma A.3.5), we have:

$$(\mathbb{E}\|\tilde{\alpha} - \tilde{\beta}\|_E^2)^{1/2} \leq (1 + HL)e_{N-1} \tag{A.171}$$

All in all,

$$\begin{aligned} e_N &\leq (\mathbb{E}\|x(NH) - \tilde{\alpha}\|_E^2)^{1/2} + (\mathbb{E}\|\tilde{\alpha} - \tilde{\beta}\|_E^2)^{1/2} + (\mathbb{E}\|\tilde{\beta} - X_{NH})\|_E^2)^{1/2} \\ &\leq (1 + HL)e_{N-1} + (C_1 + C_2)H^{3/2} \\ &= (1 + HL)^N e_0 + (C_1 + C_2)H^{3/2}\frac{(1 + HL)^N - 1}{(1 + HL) - 1} \\ &\leq (C_1 + C_2)H^{1/2}\frac{e^{NHL} - 1}{L} = \frac{(C_1 + C_2)(e^{TL} - 1)}{L}H^{1/2} \end{aligned} \tag{A.172}$$

Therefore letting $C = \frac{(C_1+C_2)(e^{TL}-1)}{L}$ we have

$$(\mathbb{E}\|q(T) - q_T\|_2^2)^{1/2} \leq e_N \leq CH^{1/2} \tag{A.173}$$

$$(\mathbb{E}\|p(T) - p_T\|_2^2)^{1/2} \leq \epsilon^{-1/2}\|\sqrt{K}\|_2 e_N \leq \epsilon^{-1/2}\|\sqrt{K}\|_2 CH^{1/2} \tag{A.174}$$

□

Bibliography

[1] A. Abdulle. Fourth order Chebyshev methods with recurrence relation. *SIAM J. Sci. Comput.*, 23(6):2041–2054 (electronic), 2002.

[2] A. Abdulle and S. Cirilli. S-ROCK: Chebyshev methods for stiff stochastic differential equations. *SIAM J. Sci. Comput.*, 30(2):997–1014, 2008.

[3] Ralph Abraham and Jerrold E. Marsden. *Foundations of Mechanics*. American Mathematical Society, 2nd edition, 2008.

[4] G. Allaire. Homogenization and two-scale convergence. *SIAM J. Math. Anal.*, 23(6):1482–1518, 1992.

[5] G. Allaire. *Shape optimization by the homogenization method*, volume 146 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2002.

[6] G. Allaire and R. Brizzi. A multiscale finite element method for numerical homogenization. *Multiscale Model. Simul.*, 4(3):790–812 (electronic), 2005.

[7] Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72(4):2384–2393, 1980.

[8] Hans C Andersen. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys*, 52(1):24–34, 1983.

[9] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. SIAM, 3rd edition, 1999.

[10] H. Anderson. RATTLE: A velocity version of the SHAKE algorithm for molecular dynamics calculations. *J. Comput. Phys.*, 52:24–34, 1983.

[11] T. Arbogast and K. J. Boyd. Subgrid upscaling and mixed multiscale finite elements. *SIAM J. Numer. Anal.*, 44(3):1150–1171 (electronic), 2006.

[12] T. Arbogast, C.-S. Huang, and S.-M. Yang. Improved accuracy for alternating-direction methods for parabolic equations based on regular and mixed finite elements. *Math. Models Methods Appl. Sci.*, 17(8):1279–1305, 2007.

[13] G. Ariel, B. Engquist, and Y.-H.R. Tsai. A multiscale method for highly oscillatory ordinary differential equations with resonance. *Math. Comput.*, 78:929, 2009.

[14] Gil Ariel, Bjorn Engquist, and Richard Tsai. A reversible multiscale integration method. *Commun. Math. Sci.*, 7(3):595–610, 2009.

[15] A. Armaou and I. G. Kevrekidis. Equation-free optimal switching policies for bistable reacting systems. *Internat. J. Robust Nonlinear Control*, 15(15):713–726, 2005.

[16] S.S. Artem′ev and Kh. Shurts. *Zhestkie sistemy stokhasticheskikh differentsialnykh uravnenii s malym shumom i ikh chislennoe reshenie*, volume 1039. Ross. Akad. Nauk Sibirsk. Otdel. Vychisl. Tsentr, Novosibirsk, 1995.

[17] Z. Artstein, I. G. Kevrekidis, M. Slemrod, and E. S. Titi. Slow observables of singularly perturbed differential equations. *Nonlinearity*, 20(11):2463–2481, 2007.

[18] Z. Artstein, J. Linshiz, and E. S. Titi. Young measure approach to computing slowly advancing fast oscillations. *Multiscale Model. Simul.*, 6(4):1085–1097, 2007.

[19] Zvi Artstein, C. William Gear, Ioannis G. Kevrekidis, Marshall Slemrod, and Edriss S. Titi. Analysis and computation of a discrete kdv-burgers type equation with fast dispertion and slow diffusion. arXiv:0908.2752, 2009.

[20] I. Babuška and R. Lipton. Optimal local approximation spaces for generalized finite element methods with application to multiscale problems. *Multiscale Model. Simul.*, 9:373–406, 2011.

[21] I. Babuška and John E. Osborn. Generalized finite element methods: Their performance and their relation to mixed methods. *SIAM J. Numer. Anal.*, 20:510–536, 1983.

[22] Ivo Babuška, Gabriel Caloz, and John E. Osborn. Special finite element methods for a class of second order elliptic problems with rough coefficients. *SIAM J. Numer. Anal.*, 31(4):945–981, 1994.

[23] N. Bakhvalov and G. Panasenko. Homogenization: averaging processes in periodic media. In *Mathematics and its applications, vol. 36*. Kluwer Academic Publishers, Dordrecht, 1990.

[24] Eric J. Barth, Brian B. Laird, and Benedict J. Leimkuhler. Generating generalized distributions from dynamical simulation. *J. Chem. Phys.*, 118(13):5759–5768, 2003.

[25] Federico Bassetti and Persi Diaconis. Examples comparing importance sampling and the metropolis algorithm. *Illinois J. of Math.*, 50(1):67–91, 2005.

[26] Oren M. Becker, Alexander D. MacKerell, Jr., Benoit Roux, and Masakatsu Watanabe. *Computational Biochemistry and Biophysics*. Marcel Dekker Ltd., 2001.

[27] A. Bensoussan, J. L. Lions, and G. Papanicolaou. *Asymptotic analysis for periodic structure*. North Holland, Amsterdam, 1978.

[28] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[29] L. Berlyand and H. Owhadi. Flux norm approach to finite dimensional homogenization approximations with non-separated scales and high contrast. *Archives for Rational Mechanics and Analysis*, 198(2):677–721, 2010.

[30] C. Bernardi and R. Verfürth. Adaptive finite element methods for elliptic equations with non-smooth coefficients. *Numer. Math.*, 85(4):579–608, 2000.

[31] Bruce J Berne and John E Straub. Novel methods of sampling phase space in the simulation of biological systems. *Curr. Opin. Struct. Biol.*, 7(2):181–189, 1997.

[32] Alexandros Beskos, F.J. Pinski, J. M. Sanz-Serna, and Andrew Stuart. Hybrid Monte-Carlo on Hilbert spaces. Submitted, 2011.

[33] J.M. Bismut. *Mécanique aléatoire*. Springer, 1981.

[34] X. Blanc, C. Le Bris, and P.-L. Lions. Une variante de la théorie de l'homogénéisation stochastique des opérateurs elliptiques. *C. R. Math. Acad. Sci. Paris*, 343(11-12):717–724, 2006.

[35] X. Blanc, C. Le Bris, and P.-L. Lions. Stochastic homogenization and random lattices. *J. Math. Pures Appl. (9)*, 88(1):34–63, 2007.

[36] N.N. Bogolyubov. *Problemy dinamičeskoĭ teorii v statističeskoĭ fizike*. Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow-Leningrad, 1946.

[37] Folkmar A. Bornemann and Christof Schütte. A mathematical approach to smoothed molecular dynamics: Correcting potentials for freezing bond angles. Prprint SC 9-30 (Dember 1995), 1995.

[38] Folkmar A. Bornemann and Christof Schütte. Homogenization of Hamiltonian systems with a strong constraining potential. *Phys. D*, 102(1-2):57–77, 1997.

[39] N. Bou-Rabee and J. Marsden. Hamilton-Pontryagin integrators on Lie groups. I. Introduction and structure-preserving properties. *Found. Comput. Math.*, 9(2):197–219, 2009.

[40] N. Bou-Rabee and H. Owhadi. Stochastic variational integrators. *IMA J. Numer. Anal.*, 29(2):421–443, 2009.

[41] Nawaf Bou-Rabee and Houman Owhadi. Long-run accuracy of variational integrators in the stochastic context. *SIAM J. Numer. Anal.*, 48(1):278–297, 2010.

[42] Nawaf Bou-Rabee and Eric Vanden-Eijnden. Pathwise accuracy and ergodicity of metropolized integrators for SDEs. *Comm. Pure Appl. Math.*, 63(5):655–696, 2010.

[43] A. Braides. Γ-*convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2002.

[44] L. V. Branets, S. S. Ghai, L. L., and X.-H. Wu. Challenges and technologies in reservoir modeling. *Commun. Comput. Phys.*, 6(1):1–23, 2009.

[45] Thomas J. Bridges. A geometric formulation of the conservation of wave action and its implications for signature and the classification of instabilities. *Proc. Roy. Soc. Lond. A*, 453(1962):1365–1395, 1997.

[46] Thomas J. Bridges. Multi-symplectic structures and wave propagation. *Math. Proc. Camb. Phil. Soc.*, 121(01):147–190, 1997.

[47] Thomas J. Bridges. Canonical multi-symplectic structure on the total exterior algebra bundle. *Proc. Roy. Soc. Lond. A*, 462(2069):1531–1551, 2006.

[48] Thomas J. Bridges and Gianne Derks. Unstable eigenvalues and the linearization about solitary waves and fronts with symmetry. *Proc. Roy. Soc. Lond. A*, 455(1987):2427–2469, 1999.

[49] Thomas J. Bridges and Sebastian Reich. Multi-symplectic integrators: numerical schemes for Hamiltonian PDEs that conserve symplecticity. *Phys. Lett. A*, 284(4–5):184–193, 2001.

[50] Thomas J Bridges and Sebastian Reich. Numerical methods for Hamiltonian PDEs. *J. Phys. A*, 39(19):5287–5320, 2006.

[51] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comp. Chem.*, 4:187–217, 1983.

[52] K. Burrage and T. Tian. The composite Euler method for stiff stochastic differential equations. *J. Comput. Appl. Math.*, 131(1-2):407–426, 2001.

[53] K. Burrage and T. Tian. Stiffly accurate Runge-Kutta methods for stiff stochastic differential equations. *Comput. Phys. Comm.*, 142(1-3):186–190, 2001. Computational physics 2000. "New challenges for the new millenium" (Gold Coast).

[54] L. A. Caffarelli and P. E. Souganidis. A rate of convergence for monotone finite difference approximations to fully nonlinear, uniformly elliptic PDEs. *Comm. Pure Appl. Math.*, 61(1):1–17, 2008.

[55] M. P. Calvo and J. M. Sanz-Serna. Instabilities and inaccuracies in the integration of highly oscillatory problems. *SIAM J. Sci. Comput.*, 31(3):1653–1677, 2009.

[56] M. P. Calvo and J. M. Sanz-Serna. Heterogeneous multiscale methods for mechanical systems with vibrations. *SIAM J. Sci. Comput., to appear*, 2010.

[57] Eric Cancès, Frédéric Legoll, and Gabriel Stoltz. Theoretical and numerical comparison of some sampling methods for molecular dynamics. *M2AN*, 41(2):351–389, Mar 2007.

[58] E.J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, feb. 2006.

[59] F. Castella, P. Chartier, and E. Faou. An averaging technique for highly-oscillatory hamiltonian problems. *SIAM J. Numer. Anal.*, 47:2808–2837, 2009.

[60] P J Channell and C Scovel. Symplectic integration of hamiltonian systems. *Nonlinearity*, 3(2):231, 1990.

[61] J. G. Charney, R. Fjörtoft, and J. von Neumann. Numerical integration of the barotropic vorticity equation. *Tellus*, 2:237–254, 1950.

[62] Jing-Bo Chen. Symplectic and multisymplectic Fourier pseudospectral discretizations for the Klein–Gordon equation. *Lett. Math. Phys.*, 75(3):293–305, 2006.

[63] Sai Hung Cheung and James L. Beck. Bayesian model updating using hybrid monte carlo simulation with application to structural dynamic models with many uncertain parameters. *J. Eng. Mech.*, 135(4):243–255, 2009.

[64] Fumihiro Chiba and Takashi Kako. Newmark's method and discrete energy applied to resistive mhd equation. *Vietnam J. Math.*, 30:501–520, 2002.

[65] A.J. Chorin, O.H. Hald, and R. Kupferman. Optimal prediction and the Mori-Zwanzig representation of irreversible processes. *Proc. Natl. Acad. Sci. USA*, 97(7):2968–2973 (electronic), 2000.

[66] A.J. Chorin, O.H. Hald, and R. Kupferman. Optimal prediction with memory. *Phys. D*, 166(3-4):239–257, 2002.

[67] A.J. Chorin, A.P. Kast, and R. Kupferman. Optimal prediction of underresolved dynamics. *Proc. Natl. Acad. Sci. USA*, 95(8):4094–4098 (electronic), 1998.

[68] C.-C. Chu, I. G. Graham, and T. Y. Hou. A new multiscale finite element method for high-contrast elliptic interface problems. *Math. Comp.*, 79:1915–1955, 2010.

[69] G. Ciccotti, T. Lelievre, and E. Vanden-Eijnden. Projections of diffusions on submanifolds: Application to mean force computation. *Comm. Pure Appl. Math.*, 61:371–408, 2008.

[70] D. Cioranescu and P. Donato. *An introduction to homogenization*, volume 17 of *Oxford Lecture Series in Mathematics and its Applications*. The Clarendon Press Oxford University Press, New York, 1999.

[71] D. Cohen, T. Jahnke, K. Lorenz, and C. Lubich. Numerical integrators for highly oscillatory Hamiltonian systems: a review. In *Analysis, modeling and simulation of multiscale problems*, pages 553–576. Springer, Berlin, 2006.

[72] David Cohen, Ernst Hairer, and Christian Lubich. Modulated fourier expansions of highly oscillatory differential equations. *Foundations of Computational Mathematics*, 3:327–345, 2003. 10.1007/s10208-002-0062-x.

[73] Harry Cohn and Mark Fielding. Simulated annealing: Searching for an optimal temperature schedule. *SIAM J. Optimiz.*, 9(3):779–802, 1999.

[74] M. Condon, A. Deaño, and A. Iserles. On highly oscillatory problems arising in electronic engineering. *M2AN Math. Model. Numer. Anal.*, 43(4):785–804, 2009.

[75] Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progressions. *J. Symbolic Comput.*, 9(3):251–280, 1990. Computational algebraic complexity editorial.

[76] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M.Jr Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[77] R Courant, K Friedrichs, and H Lewy. Über die partiellen Differenzengleichungen der mathematischen Physik. *Math. Ann.*, 100:32–74, 1928.

[78] Felipe Cucker and Antonio Gonzalez Corbalan. An alternate proof of the continuity of the roots of a polynomial. *Amer. Math. Monthly.*, 96(4):342–345, 1989.

[79] G. Dahlquist. Convergence and stability in the numerical integration of ordinary differential equations. *Math. Scand.*, 4:33–53, 1956.

[80] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*. Springer, New York, second edition, 1998.

[81] Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov Chains. *Ann. Appl. Probab.*, 1:36–61, 1991.

[82] Matthew Dobson, Claude Le Bris, and Frederic Legoll. Symplectic schemes for highly oscillatory Hamiltonian systems: the homogenization approach beyond the constant frequency case. arXiv:1008.1030, 2010.

[83] D.L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, april 2006.

[84] Simon Duane, A.D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid monte carlo. *Phys. Lett. B*, 195(2):216–222, 1987.

[85] W. E. Analysis of the heterogeneous multiscale method for ordinary differential equations. *Commun. Math. Sci.*, 1(3):423–436, 2003.

[86] W. E, B. Engquist, X. Li, W. Ren, and E. Vanden-Eijnden. Heterogeneous multiscale methods: a review. *Commun. Comput. Phys.*, 2(3):367–450, 2007.

[87] W. E, D. Liu, and E. Vanden-Eijnden. Analysis of multiscale methods for stochastic differential equations. *Comm. Pure Appl. Math.*, 58(11):1544–1585, 2005.

[88] W. E, D. Liu, and E. Vanden-Eijnden. Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales. *J. Comput. Phys.*, 221(1):158–180, 2007.

[89] W. E, W. Ren, and E. Vanden-Eijnden. A general strategy for designing seamless multiscale methods. *J. Comput. Phys.*, 228:5437–5453, 2009.

[90] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. *Phys. Rev. B*, 66:052301, 2002.

[91] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.*, 126:164103, 2007.

[92] Y. Efendiev, J. Galvis, and X. Wu. Multiscale finite element and domain decomposition methods for high-contrast problems using local spectral basis functions. 2009. Submitted.

[93] Yalchin Efendiev and Thomas Y. Hou. *Multiscale Finite Element Methods*. Springer, 1st edition, 2009.

[94] Yalchin Efendiev, Thomas Y. Hou, and Victor Ginting. Multiscale finite element methods for nonlinear problems and their applications. *Comm. Math. Sci.*, 2:553–589, 2004.

[95] Stanley C. Eisenstat and Ilse C. F. Ipsen. Relative perturbation techniques for singular value problems. *SIAM J. Num. Anal.*, 32:1972, 1995.

[96] B. Engquist and P. E. Souganidis. Asymptotic and numerical homogenization. *Acta Numerica*, 17:147–190, 2008.

[97] B. Engquist and Y.-H.R. Tsai. Heterogeneous multiscale methods for stiff ordinary differential equations. *Math. Comp.*, 74(252):1707–1742 (electronic), 2005.

[98] S. Erlicher, L. Bonaventura, and O. S. Bursi. The analysis of the Generalized α method for non-linear dynamic problems. *Comput. Mech.*, 28(2):83–104, 2002.

[99] Lawrence C. Evans. *Partial Differential Equations*. American Mathematical Society, 1998.

[100] L.C. Evans. A survey of partial differential equations methods in weak KAM theory. *Comm. Pure Appl. Math.*, 57(4):445–480, 2004.

[101] E. Fermi, J. Pasta, and S. Ulam. Studies of nonlinear problems. Technical Report LA-1940, Los Alamos Scientific Laboratory, 1955.

[102] Francis Filbet and Shi Jin. A class of asymptotic preserving schemes for kinetic equations and related problems with stiff sources. 2010. arXiv:0905.1378. Accepted by J. Comput. Phys.

[103] M. Fixman. Classical statistical mechanics of constraints: A theorem and application to polymers. *Proc. Nat. Acad. Sci. USA*, 71-8:3050–3053, 1974.

[104] S. Flach, M. V. Ivanchenko, and O. I. Kanakov. q-breathers and the fermi-pasta-ulam problem. *Phys. Rev. Lett.*, 95(6):064102, Aug 2005.

[105] Gaston Floquet. Sur les équations différentielles linéaires à coefficients périodiques. *Ann. École Norm. Sup.*, 12:47–88, 1883.

[106] Joseph Ford. The fermi-pasta-ulam problem: Paradox turns discovery. *Phys. Rep.*, 213(5):271 – 310, 1992.

[107] M.I. Freidlin and A.D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, New York USA, second edition, 1998.

[108] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Academic Press, Orlando USA, second edition, 2002.

[109] B. García-Archilla, J. M. Sanz-Serna, and R. D. Skeel. Long-time-step methods for oscillatory differential equations. *SIAM J. Sci. Comput.*, 20(3):930–963, 1999.

[110] W. Gautschi. Numerical integration of ordinary differential equations based on trigonometric polynomials. *Numer. Math.*, 3:381–397, 1961.

[111] C. W. Gear and I. G. Kevrekidis. Projective methods for stiff differential equations: problems with gaps in their eigenvalue spectrum. *SIAM J. Sci. Comput.*, 24(4):1091–1106 (electronic), 2003.

[112] C.W. Gear. *Numerical initial value problems in ordinary differential equations*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1971.

[113] C.W. Gear and K.A. Gallivan. Automatic methods for highly oscillatory ordinary differential equations. In *Numerical analysis (Dundee, 1981)*, volume 912 of *Lecture Notes in Math.*, pages 115–124. Springer, Berlin, 1982.

[114] Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(6):721–741, November 1984.

[115] I.I. Gihman. On the theory of differential equations of stochastic processes. I, II. *Amer. Math. Soc. Transl. (2)*, 1:111–137, 139–161, 1955.

[116] E. De Giorgi. Sulla convergenza di alcune successioni di integrali del tipo dell'aera. *Rendi Conti di Mat.*, 8:277–294, 1975.

[117] E. De Giorgi. New problems in Γ-convergence and G-convergence. In *Free boundary problems, Vol. II (Pavia, 1979)*, pages 183–194. Ist. Naz. Alta Mat. Francesco Severi, Rome, 1980.

[118] D. Givon, I. G. Kevrekidis, and R. Kupferman. Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems. *Commun. Math. Sci.*, 4(4):707–729, 2006.

[119] A. Gloria. Analytical framework for the numerical homogenization of elliptic monotone operators and quasiconvex energies. *SIAM MMS*, 5(3):996–1043, 2006.

[120] Sergei Konstantinovich Godunov. A difference scheme for numerical computation of discontinuous solutions of hydrodynamic equations. *Math. Sbornik*, 47:271–306, 1959.

[121] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1996.

[122] Jonathan Goodman and Jonathan Weare. Ensemble samplers with affine invariance. *Comm. App. Math. Comp. Sci.*, 5(1):65–80, 2010.

[123] Volker Grimm and Marlis Hochbruck. Error analysis of exponential integrators for oscillatory second-order differential equations. *J. Phys. A: Math. Gen.*, 39:5495–5507, 2006.

[124] H. Grubmuller, H. Heller, A. Windemuth, and K. Schulten. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Simul.*, 6:121–142, 1991.

[125] Helmut Grubmüller. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E*, 52(3):2893–2906, Sep 1995.

[126] S.A. Gusev. *Algoritm peremennogo shaga dlya chislennogo resheniya zhestkikh sistem stokhasticheskikh differentsialnykh uravnenii*, volume 1094. Rossiĭskaya Akademiya Nauk Sibirskoe Otdelenie, Institut Vychislitel′noĭ Matematiki i Matematicheskoĭ Geofiziki, Novosibirsk, 1997.

[127] Olle Häggström. *Finite Markov Chains and Algorithmic Applications*. Cambridge University Press, Cambridge UK, first edition, 2002.

[128] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer, Heidelberg Germany, second edition, 2004.

[129] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving ordinary differential equations. I*, volume 8 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 1993. Nonstiff problems.

[130] E. Hairer and G. Wanner. *Solving ordinary differential equations. II*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 1996. Stiff and differential-algebraic problems.

[131] Ernst Hairer, Christian Lubich, and Gerhard Wanner. Geometric numerical integration illustrated by the Störmer-Verlet method. *Acta Numer.*, 12:399–450, 2003.

[132] Bruce Hajek. Cooling schedules for optimal annealing. *Math. Oper. Res.*, 13(2):311–329, 1988.

[133] Hadi Hajibeygi and Patrick Jenny. Multiscale finite-volume method for parabolic problems arising from compressible multiphase flow in porous media. *J. Comput. Phys.*, 228(14):5129–5147, 2009.

[134] Donald Hamelberg, John Mongan, and J. Andrew McCammon. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.*, 120(24):11919–11929, 2004.

[135] H. Harbrecht, R. Schneider, and C. Schwab. Sparse second moment analysis for elliptic problems in stochastic domains. *Numer. Math.*, 109(3):385–414, 2008.

[136] C. Hartmann. An ergodic sampling scheme for constrained Hamiltonian systems with applications to molecular dynamics. *J. Stat. Phys.*, 130(4):687–711, 2008.

[137] Babak Hassibi. An efficient square-root algorithm for blast. In *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 737–740, Washington DC, 2000. IEEE Computer Society.

[138] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[139] Jan S. Hesthaven, Sigal Gottlieb, and David Gottlieb. *Spectral Methods for Time-Dependent Problems*, volume 21 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, United Kingdom, 2007.

[140] Nicholas J. Higham. Stable iterations for the matrix square root. *Numer. Algorithms*, 15(2):227–242, 1997.

[141] Nicholas J. Higham. The scaling and squaring method for the matrix exponential revisited. *SIAM J. Matrix Anal. Appl.*, 26:1179–1193, 2005.

[142] George William Hill. On the part of the motion of lunar perigee which is a function of the mean motions of the sun and moon. *Acta Math.*, 8(1):1–36, 1886.

[143] Marlis Hochbruck and Christian Lubich. A gautschi-type method for oscillatory second-order differential equations. *Numer. Math.*, 83:403–426, 1999.

[144] William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, Mar 1985.

[145] RA Horn and CR Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[146] Thomas Y. Hou and Xiao-Hui Wu. A multiscale finite element method for elliptic problems in composite materials and porous media. *J. Comput. Phys.*, 134(1):169 – 189, 1997.

[147] Thomas Y. Hou, Xiao-Hui Wu, and Zhiqiang Cai. Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients. *Math. Comp.*

[148] Thomas Y. Hou, Danping Yang, and Hongyu Ran. Multiscale analysis and computation for the three-dimensional incompressible naiver stokes. *Multiscale Model. Simul.*, 6(4):1317–1346, 2008.

[149] Thomas J. R. Hughes. A note on the stability of newmark's algorithm in nonlinear structural dynamics. *Int. J. Numer. Meth. Eng.*, 11(2):383–386, 1977.

[150] A. Iserles. On the global error of discretization methods for highly-oscillatory ordinary differential equations. *BIT*, 42(3):561–599, 2002.

[151] A. Iserles. Think globally, act locally: solving highly-oscillatory ordinary differential equations. *Appl. Numer. Math.*, 43(1-2):145–160, 2002. 19th Dundee Biennial Conference on Numerical Analysis (2001).

[152] A. Iserles. On the numerical analysis of rapid oscillation. In *Group theory and numerical analysis*, volume 39 of *CRM Proc. Lecture Notes*, pages 149–163. Amer. Math. Soc., Providence, RI, 2005.

[153] A. Iserles, S. P. Nørsett, and S. Olver. Highly oscillatory quadrature: the story so far. In *Numerical mathematics and advanced applications*, pages 97–118. Springer, Berlin, 2006.

[154] Arieh Iserles, Hans Z. Munthe-Kaas, Syvert P. NØrsett, and Antonella Zanna. Lie-group methods. *Acta Numerica*, 9:215–365, 2000.

[155] A. Jain, N. Vaidehi, and G. Rodriguez. A fast recursive algorithm for molecular dynamics simulation. *J. Comput. Phys*, 106(2):258–268, 1993.

[156] P. Jenny, S. H. Lee, and H. A. Tchelepi. Multi-scale finite-volume method for elliptic problems in subsurface flow simulation. *J. Comput. Phys.*, 187(1):47–67, 2003.

[157] V. V. Jikov, S. M. Kozlov, and O. A. Oleĭnik. *Homogenization of differential operators and integral functionals*. Springer-Verlag, Berlin, 1994.

[158] V. V. Jikov, S. M. Kozlov, and O. A. Oleinik. *Homogenization of Differential Operators and Integral Functionals*. Springer-Verlag, Berlin, 1994.

[159] Shi Jin, Lorenzo Pareschi, and Giuseppe Toscani. Uniformly accurate diffusive relaxation schemes for multiscale transport equations. *SIAM J. Numer. Anal.*, 38(3):913–936 (electronic), 2000.

[160] Yun Jing, Molei Tao, and Gregory Clement. Evaluation of a wave vector frequency domain method for nonlinear wave propagation. *J. Acoust. Soc. Am.*, 129(1):32–46, 2011.

[161] Couro Kane, Jerrold E Marsden, Michael Ortiz, and Matthew West. Variational integrators and the Newmark algorithm for conservative and dissipative mechanical systems. *Int. J. Numer. Meth. Eng.*, 49:1295–1325, 2000.

[162] P.L. Kapitza. *Collected Papers of P.L.Kapitza, Volume II., edited by D. Ter Haar*. Pergamon Press, Oxford UK, 1965.

[163] Frithjof Karsch and Edwin Laermann. Numerical simulations in particle physics. *Rep. Prog. Phys.*, 56:1347–1395, 1993.

[164] I. Kevrekidis and G. Samaey. Equation-free multiscale computation: Algorithms and applications. *Annual Review of Physical Chemistry*, 60(1):321–344, 2009. PMID: 19335220.

[165] I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg, and C. Theodoropoulos. Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis. *Commun. Math. Sci.*, 1(4):715–762, 2003.

[166] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[167] S. M. Kozlov. The averaging of random operators. *Mat. Sb. (N.S.)*, 109(151)(2):188–202, 327, 1979.

[168] H.-O. Kreiss. Problems with different time scales. *Acta Numer.*, 1:101–139, 1992.

[169] B. Kryloff and N. Bogoliouboff. La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire. *Ann. of Math. (2)*, 38(1):65–113, 1937.

[170] N. Kryloff and N. Bogoliouboff. On some problems in the ergodic theory of stochastic systems. *Zap. Kafedr. Mat. Fiz. Inst. Budivel. Mat. Akad. Nauk. Ukrain. SSR*, 4:243–287, 1939.

[171] D. Kuhl and M. A. Crisfield. Energy-conserving and decaying algorithms in non-linear structural dynamics. *Int. J. Numer. Meth. Engng.*, 45(5):569–599, 1999.

[172] Detlef Kuhl and Ekkehard Ramm. Constraint energy momentum algorithm and its application to non-linear dynamics of shells. *Computer Methods in Applied Mechanics and Engineering*, 136(3-4):293 – 315, 1996.

[173] Keith J. Laidler and M. Christine King. Development of transition-state theory. *J. Phys. Chem.*, 87(15):2657–2664, 1983.

[174] P.D. Lax and R.D. Richtmyer. Survey of the stability of linear finite difference equations. *Comm. Pure Appl. Math.*, 9:267–293, 1956.

[175] J.-A. Lázaro-Camí and J. P. Ortega. Stochastic Hamiltonian dynamical systems. *Rep. Math. Phys.*, 61(1):65–122, 2008.

[176] Claude Le Bris and Frédéric Legoll. Integrators for highly oscillatory hamiltonian systems: An homogenization approach. *Discrete Contin. Dyn. Syst. Ser. B*, 13:347–373, 2010.

[177] V.I. Lebedev and S.A. Finogenov. The use of ordered Čebyšev parameters in iteration methods. *Ž. Vyčisl. Mat. i Mat. Fiz.*, 16(4):895–907, 1084, 1976.

[178] B. Leimkuhler and S. Reich. A reversible averaging integrator for multiple time-scale dynamics. *J. Comput. Phys.*, 171(1):95–114, 2001.

[179] B. Leimkuhler and S. Reich. *Simulating Hamiltonian dynamics*, volume 14 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2004.

[180] B. J. Leimkuhler, S. Reich, and R. D. Skeel. Integration methods for molecular dynamics. In *Mathematical approaches to biomolecular structure and dynamics (Minneapolis, MN, 1994)*, volume 82 of *IMA Vol. Math. Appl.*, pages 161–185. Springer, New York, 1996.

[181] Ben Leimkuhler and Sebastian Reich. A reversible averaging integrator for multiple time-scale dynamics. *J. Comput. Phys.*, 171:95–114, 2001.

[182] Randall J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 1st edition, 2002.

[183] A. Lew, J.E. Marsden, M. Ortiz, and M. West. Asynchronous variational integrators. *Arch. Ration. Mech. Anal.*, 167:85–146, 2003.

[184] T. Li, A. Abdulle, and W. E. Effectiveness of implicit methods for stiff stochastic differential equations. *Commun. Comput. Phys.*, 3(2):295–307, 2008.

[185] Stefano Luzzatto, Ian Melbourne, and Frederic Paccaut. The lorenz attractor is mixing. *Comm. publisher = Springer Berlin / Heidelberg, issn = 0010-3616, keyword = Physics and Astronomy, pages = 393-401, volume*

= 260, issue = 2, url = http://dx.doi.org/10.1007/s00220-005-1411-9, note = 10.1007/s00220-005-1411-9, year = 2005.

[186] R S MacKay and S Aubry. Proof of existence of breathers for time-reversible or hamiltonian networks of weakly coupled oscillators. *Nonlinearity*, 7(6):1623, 1994.

[187] Wilhelm Magnus. On the exponential solution of differential equations for a linear operator. *Comm. Pure Appl. Math.*, 7:649–673, 1954.

[188] S. J. A. Malham and A. Wiese. Stochastic Lie group integrators. *SIAM J. Sci. Comput.*, 30(2):597–617, 2008.

[189] Luca Maragliano and Eric Vanden-Eijnden. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem. Phys. Lett.*, 426(1-3):168 – 175, 2006.

[190] Luca Maragliano and Eric Vanden-Eijnden. On-the-fly string method for minimum free energy paths calculation. *Chem. Phys. Lett.*, 446(1-3):182–190, 2007.

[191] Gurii Ivanovich Marchuk. Some applications of splitting-up methods to the solution of mathematical physics problems. *Applik. Mat.*, 13:103–132, 1968.

[192] J. E. Marsden and M. West. Discrete mechanics and variational integrators. *Acta Numerica*, pages 357–514, 2001.

[193] Jerrold E. Marsden, George W. Patrick, and Steve Shkoller. Multisymplectic geometry, variational integrators, and nonlinear PDEs. *Commun. Math. Phys.*, 199(2):351–395, 1998.

[194] Jerrold E Marsden and Tudor S Ratiu. *Introduction to Mechanics and Symmetry*. Springer, 2nd edition, 2010.

[195] Jerrold E. Marsden and Steve Shkoller. Multisymplectic geometry, covariant

hamiltonians, and water waves. *Math. Proc. Camb. Phil. Soc.*, 125(03):553–575, 1999.

[196] Emile Mathieu. Mémoire sur le mouvement vibratoire d'une membrane de forme elliptique. *J. Math. Pures Appl.*, 13:137–203, 1868.

[197] J. C. Mattingly and A. M. Stuart. Geometric ergodicity of some hypo-elliptic diffusions for particle motions. *Markov Process. Related Fields*, 8(2):199–214, 2002. Inhomogeneous random systems (Cergy-Pontoise, 2001).

[198] J.C. Mattingly, A.M. Stuart, and D.J. Higham. Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. *Stoch. Proc. Appl.*, 101:185–232, 2002.

[199] R. McLachlan, G. Reinout, and W. Quispel. Splitting methods. *Acta Numerica*, pages 341–434, 2002.

[200] R I McLachlan and P Atela. The accuracy of symplectic integrators. *Nonlinearity*, 5(2):541, 1992.

[201] R. I. McLachlan and D. R. J. OíNeale. Comparison of integrators for the Fermi-Pasta-Ulam problem. *preprint NI07052-HOP, Isaac Newton Institute for Mathematical Sciences.*, 2007. http://www.newton.ac.uk/preprints/NI07052.pdf.

[202] R. I. McLachlan and M. Perlmutter. Conformal Hamiltonian systems. *J. Geom. Phys.*, 39(4):276–300, 2001.

[203] G. N. Milstein, Yu. M. Repin, and M. V. Tretyakov. Symplectic integration of Hamiltonian systems with additive noise. *SIAM J. Numer. Anal.*, 39(6):2066–2088, 2002.

[204] G. N. Milstein, Yu. M. Repin, and M. V. Tretyakov. Numerical methods for stochastic systems preserving symplectic structure. *SIAM J. Numer. Anal.*, 40(4):1583–1604, 2003.

[205] G. N. Milstein and M. V. Tretyakov. Quasi-symplectic methods for Langevin-type equations. *IMA J. Numer. Anal.*, 23(4):593–626, 2003.

[206] G. N. Milstein and M. V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Springer, 2004.

[207] Shuichi Miyamoto and Peter A. Kollman. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.*, 13(8):952–962, 1992.

[208] Cleve Moler and Charles Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, 45:3–49, 2003.

[209] F. Murat and L. Tartar. H-convergence. *Séminaire d'Analyse Fonctionnelle et Numérique de l'Université d'Alger*, 1978.

[210] François Murat. Compacité par compensation. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 5(3):489–507, 1978.

[211] Giovanni Naldi and Lorenzo Pareschi. Numerical schemes for hyperbolic systems of conservation laws with stiff diffusive relaxation. *SIAM J. Numer. Anal.*, 37(4):1246–1270 (electronic), 2000.

[212] Radford M. Neal. Sampling from multimodal distributions using tempered transitions. *Statist. Comput.*, 6(4):353–366, 1996.

[213] F. Neri. Lie algebras and canonical integration. Technical report, Department of Physics, University of Maryland, 1988.

[214] N. M. Newmark. A method of computation for structural dynamics. *Proc. ASCE*, 85(3):67–94, 1959.

[215] G. Nguetseng. A general convergence result for a functional related to the theory of homogenization. *SIAM J. Math. Anal.*, 20(3):608–623, 1989.

[216] J. Nolen, G. Papanicolaou, and O. Pironneau. A framework for adaptive multiscale methods for elliptic problems. *Multiscale Model. Simul.*, 7(1):171–196, 2008.

[217] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 52:255–268, 1984.

[218] Sina Ober-Blöbaum, Molei Tao, Mulin Cheng, Houman Owhadi, and Jerrold E. Marsden. Variational integrators for electric circuits. arXiv:1103.1859. Submitted, 2011.

[219] Bernt Karsten Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, 5th edition, 2002.

[220] H. Owhadi and L. Zhang. Metric-based upscaling. *Comm. Pure Appl. Math.*, 60(5):675–723, 2007.

[221] H. Owhadi and L. Zhang. Localized bases for finite dimensional homogenization approximations with non-separated scales and high-contrast. 2010. arXiv:1011.0986.

[222] G.C. Papanicolaou and W. Kohler. Asymptotic theory of mixing stochastic ordinary differential equations. *Comm. Pure Appl. Math.*, 27:641–668, 1974.

[223] George C. Papanicolaou and S. R. S. Varadhan. Diffusions with random coefficients. In *Statistics and probability: essays in honor of C. R. Rao*, pages 547–552. North-Holland, Amsterdam, 1982.

[224] E. Pardoux and A. Yu. Veretennikov. On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29(3):pp. 1061–1085, 2001.

[225] E. Pardoux and A. Yu. Veretennikov. On Poisson equation and diffusion approximation 2. *Ann. Probab.*, 31(3):pp. 1166–1192, 2003.

[226] Grigorios A. Pavliotis and Andrew M. Stuart. *Multiscale methods*, volume 53

of *Texts in Applied Mathematics*. Springer, New York, 2008. Averaging and homogenization.

[227] Dennis Perchak, J. Skolnick, and Robert Yaris. Dynamics of rigid and flexible constraints for polymers. Effect of the Fixman potential. *Macromolecules*, 18(3):519–525, 1985.

[228] L Perko. *Differential equations and dynamical systems*. Springer, 2001.

[229] L.R. Petzold, L.O. Jay, and J. Yen. Problems with different time scales. *Acta Numer.*, 6:437–483, 1997.

[230] John C. Platt and Alan H. Barr. Constraints methods for flexible models. *SIGGRAPH Comput. Graph.*, 22:279–288, June 1988.

[231] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1992.

[232] Sebastian Reich. Smoothed langevin dynamics of highly oscillatory systems. *Phys. D*, 138(3-4):210–224, 2000.

[233] P. H. Richter and H.-J. Scholz. *Stochastic phenomena and chaotic behaviour in complex systems*. Springer-Verlag, Berlin and New York, 1984.

[234] Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[235] R.O. Rosenberg, B.J. Berne, and D. Chandler. Isomerization dynamics in liquids by molecular dynamics. *Chem. Phys. Lett.*, 75:162, 1980.

[236] J. Ryckaert, G. Ciccotti, and H. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23:327–341, 1977.

[237] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys*, 23(3):327–341, 1977.

[238] J.P. Ryckaert and A. Bellemans. Molecular dynamics of liquid n-butane near its boiling point. *Chem. Phys. Lett.*, 30:123, 1975.

[239] J. A. Sanders and F. Verhulst. *Averaging methods in nonlinear dynamical systems*, volume 59 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1985.

[240] J. M. Sanz-Serna. Mollified impulse methods for highly oscillatory differential equations. *SIAM J. Numer. Anal.*, 46 (2):1040–1059, 2008.

[241] J. M. Sanz-Serna. Stabilizing with a hammer. *Stoch. Dyn.*, 8(1):47–57, 2008.

[242] J. M. Sanz-Serna. Modulated Fourier expansions and heterogeneous multiscale methods. *IMA J. Numer. Anal.*, 29(3):595–605, 2009.

[243] J. M. Sanz-Serna, G. Ariel, and Y.-H.R. Tsai. Multiscale methods for stiff and constrained mechanical systems. *preprint*, 2009.

[244] J. M. Sanz-Serna, Gil Ariel, and Richard Tsai. Multiscale methods for stiff and constrained mechanical systems. 2010. Submitted to J. Comput. Phys.; preprint at ftp://ftp.math.ucla.edu/pub/camreport/cam09-08.pdf.

[245] Jesus M. Sanz-Serna. Symplectic integrators for hamiltonian problems: an overview. *Acta Numer.*, 1:243–286, 1992.

[246] R.E. Scheid. The accurate numerical solution of highly oscillatory ordinary differential equations. *Math. Comp.*, 41(164):487–509, 1983.

[247] Tamar Schlick. *Molecular Modeling and Simulation*. Springer, New York, 2002.

[248] T. Schneider and E. Stoll. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B*, 17(3):1302–1322, Feb 1978.

[249] Christof Schütte and Folkmar A. Bornemann. Homogenization approach to smoothed molecular dynamics. In *Proceedings of the Second World Congress of Nonlinear Analysts, Part 3 (Athens, 1996)*, volume 30, pages 1805–1814, 1997.

[250] Christof Schütte and W. Huisinga. On conformational dynamics induced by Langevin processes. In *EQUADIFF 99-International Conference on Differential Equations (B. Fiedler, K. Griger and J. Sprekels, eds.)*, volume 2, pages 1247–1262, Singapore, 2000. World Scientific.

[251] Claude E. Shannon. Communication in the presence of noise. *Proc. Inst. Radio Eng.*, 37(1):10–21, 1949.

[252] R. Sharp, Y.-H.R. Tsai, and B. Engquist. Multiple time scale numerical methods for the inverted pendulum problem. In *Multiscale methods in science and engineering*, volume 44 of *Lect. Notes Comput. Sci. Eng.*, pages 241–261. Springer, Berlin, 2005.

[253] David E. Shaw, Ron O. Dror, John K. Salmon, J. P. Grossman, Kenneth M. Mackenzie, Joseph A. Bank, Cliff Young, Martin M. Deneroff, Brannon Batson, Kevin J. Bowers, Edmond Chow, Michael P. Eastwood, Douglas J. Ierardi, John L. Klepeis, Jeffrey S. Kuskin, Richard H. Larson, Kresten Lindorff-Larsen, Paul Maragakis, Mark A. Moraes, Stefano Piana, Yibing Shan, and Brian Towles. Millisecond-scale molecular dynamics simulations on anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, SC '09, pages 39:1–39:11, New York, NY, USA, 2009. ACM.

[254] J. C. Simo and N. Tarnow. The discrete energy-momentum method. conserving algorithms for nonlinear elastodynamics. *ZAMP*, 43:757–792.

[255] R. D. Skeel and J. A. Izaguirre. An impulse integrator for Langevin dynamics. *Mol. Phys.*, 100:3885–3891, 2002.

[256] Robert D. Skeel and K. Srinivas. Nonlinear stability analysis of area-preserving integrators. *SIAM J. Numer. Anal.*, 38(1):129–148, 2000.

[257] Robert D. Skeel, Guihua Zhang, and Tamar Schlick. A family of symplectic integrators: Stability, accuracy, and molecular dynamics applications. *SIAM J. Sci. Comput.*, 18(1):203–222, 1997.

[258] A.V. Skorokhod. *Asymptotic methods in the theory of stochastic differential equations*, volume 78 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1989. Translated from the Russian by H. H. McFaden.

[259] Gerard L. G. Sleijpen and Henk A. Van der Vorsty. A Jacobi-Davidson iteration method for linear eigenvalue problems. *SIAM Review*, 42:267–293, 2000.

[260] Mads R. Sørensen and Arthur F. Voter. Temperature-accelerated dynamics for simulation of infrequent events. *J. Chem. Phys.*, 112:9599–9606, 2000.

[261] S. Spagnolo. Sulla convergenza di soluzioni di equazioni paraboliche ed ellittiche. *Ann. Scuola Norm. Sup. Pisa (3) 22 (1968), 571-597; errata, ibid. (3)*, 22:673, 1968.

[262] S. Spagnolo. Convergence in energy for elliptic operators. In *Numerical solutions of partial differential equations III Synspade 1975*. BAcademic Press New York, 1976.

[263] Volker Springel, Simon D. M. White, Adrian Jenkins, Carlos S. Frenk, Naoki Yoshida, Liang Gao, Julio Navarro, Robert Thacker, Darren Croton, John Helly, John A. Peacock, Shaun Cole, Peter Thomas, Hugh Couchman, August Evrard, Jorg Colberg, and Frazer Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, 2005.

[264] Mark Srednicki. *Quantum Field Theory*. Cambridge University Press, 1st edition, 2007.

[265] Ari Stern and Eitan Grinspun. Implicit-explicit variational integration of highly oscillatory problems. *Multiscale Model. Simul.*, 7:1779–1794, 2009.

[266] Gilbert Strang. On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.*, 5:506–517, 1968.

[267] W. B. Streett, D. J. Tildesley, and G. Saville. Multiple time-step methods in molecular dynamics. *Mol. Phys.*, 35:639–648, 1978.

[268] John C. Strikwerda. *Finite difference schemes and partial differential equations*. Society for Industrial and Applied Mathematics, 2nd edition, 2004.

[269] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314(1–2):141–151, 1999.

[270] Harold Szu and R. Hartley. Fast simulated annealing. *Phys. Lett. B*, 122:157–162, 1987.

[271] F. Takens. *Motion under the influence of a strong constraining force, in Global Theory of Dynamical Systems. edited by Z. Nitecki and C. Robinson.* Springer-Verlag, Berlin-Heidelberg Germany, 1980.

[272] Denis Talay. Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics and Stochastic Reports*, 29:1045–1129, 1990.

[273] Molei Tao, Houman Owhadi, and Jerrold E Marsden. From efficient symplectic exponentiation of matrices to symplectic integration of high-dimensional Hamiltonian systems with slowly varying quadratic stiff potentials. arXiv:1006.4659. Submitted, 2010.

[274] Molei Tao, Houman Owhadi, and Jerrold E Marsden. Nonintrusive and structure preserving multiscale integration of stiff ODEs, SDEs and Hamiltonian

systems with hidden slow dynamics via flow averaging. *Multiscale Model. Simul.*, 8:1269–1324, 2010.

[275] Molei Tao, Houman Owhadi, and Jerrold E Marsden. Nonintrusive and structure preserving multiscale integration of stiff ODEs, SDEs and Hamiltonian systems with hidden slow dynamics via flow averaging. *Multiscale Model. Simul.*, 8:1269–1324, 2010.

[276] Molei Tao, Houman Owhadi, and Jerrold E Marsden. Structure preserving Stochastic Impulse Method for stiff langevin systems with a uniform global error of order 1 or 1/2 on position. arXiv:1006.4657. Submitted, 2010.

[277] Molei Tao, Houman Owhadi, and Jerrold E. Marsden. Temperature and friction accelerated sampling of boltzmann-gibbs distribution. arXiv:1007.0995. Submitted, 2010.

[278] Molei Tao, Houman Owhadi, and Jerrold E Marsden. Space-time FLAVORS: finite difference, multisymlectic, and pseudospectral integrators for multiscale PDEs. *Dynam. Part. Differ. Eq.*, 8(1):21–46, 2011.

[279] Molei Tao, Houman Owhadi, and Jerrold E. Marsden. Stable, linearly-implicit and symplectic integrators, and fast long time simulations of constrained dynamics. arXiv:1103.4645. Submitted, 2011.

[280] T. Tian and K. Burrage. Implicit Taylor methods for stiff stochastic differential equations. *Appl. Numer. Math.*, 38(1-2):167–185, 2001.

[281] R. A. Todor and C. Schwab. Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA J. Numer. Anal.*, 27(2):232–261, 2007.

[282] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23(2):187–199, 1977.

[283] Lloyd Nicholas Trefethen and David Bau III. *Numerical linear algebra*. Society for Industrial and Applied Mathematics, Philadelphia, 1997.

[284] E. Triki, Y. Collette, and P. Siarry. A theoretical study on the behavior of simulated annealing leading to a new cooling schedule. *Eur. J. Oper. Res.*, 166:77–92, 2005.

[285] H. F. Trotter. Product of semigroups of operators. *Proc. Amer. Math. Soc.*, 10:545–551, 1959.

[286] M. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, 97:1990–2001, 1992.

[287] Pierre Turq, Frederic Lantelme, and Harold L. Friedman. Brownian dynamics: Its application to ionic solutions. *J. Chem. Phys.*, 66(7):3039–3044, 1977.

[288] Charles Van Loan. Computing integrals involving the matrix exponential. *IEEE Trans. Autom. Control*, 23(3):395–404, June 1978.

[289] E. Vanden-Eijnden. On HMM-like integrators and projective integration methods for systems with multiple time scales. *Commun. Math. Sci.*, 5(2):495–505, 2007.

[290] E. Vanden-Eijnden and G. Ciccotti. Second-order integrators for Langevin equations with holonomic constraints. *Chem. Phys. Lett.*, 429:310–316, 2006.

[291] F. Verhulst. *Nonlinear Differential Equations and Dynamical Systems*. Springer, Berlin-Heidelberg Germany, second edition, 1996.

[292] L. Verlet. Computer "experiments" on classical fluids. I. thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159(1):98+, July 1967.

[293] D. Viswanath. Global errors of numerical ODE solvers and Lyapunov's theory of stability. *IMA J. Numer. Anal.*, 21(1):387–406, 2001.

[294] Arthur F. Voter. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.*, 78(20):3908–3911, May 1997.

[295] Jeffrey M. Wendlandt and Jerrold E. Marsden. Mechanical integrators derived from a discrete variational principle. *Phys. D*, 106(3–4):223–246, 1997.

[296] C. D. White and R. N. Horne. Computing absolute transmissibility in the presence of finescale heterogeneity. *SPE Symposium on Reservoir Simulation*, page 16011, 1987.

[297] Jack Wisdom. The origin of the Kirkwood gaps - a mapping for asteroidal motion near the 3/1 commensurability. *AJ*, 87(3):577–593, 1982.

[298] W. L. Wood and M. E. Oduor. Stability properties of some algorithms for the solution of nonlinear dynamic vibration equations. *Commun. Appl. Numer. Methods*, 4(2):205–212, 1988.

[299] Xiongwu Wu and Shaomeng Wang. Enhancing systematic motion in molecular dynamics simulation. *J. Chem. Phys.*, 110(19):9401–9410, 1999.

[300] Tomohiro Yanao, Wang S. Koon, and Jerrold E. Marsden. Mass effects and internal space geometry in triatomic reaction dynamics. *Phys. Rev. A*, 73(5):052704, May 2006.

[301] Tomohiro Yanao, Wang Sang Koon, and Jerrold E. Marsden. Intramolecular energy transfer and the driving mechanisms for large-amplitude collective motions of clusters. *J. Chem. Phys.*, page 144111, 2009.

[302] H. Yoshimura and J. E. Marsden. Dirac structures and Lagrangian mechanics. Part I: Implicit Lagrangian systems. *J. Geom. Phys.*, 57:133–156, 2006.

[303] H. Yoshimura and J. E. Marsden. Dirac structures and Lagrangian mechanics. Part II: Variational structures. *J. Geom. Phys.*, 57:209–250, 2006.

[304] G. Zhang and T. Schlick. LIN: A new algorithm to simulate the dynamics of

biomolecules by combining implicit-integration and normal mode techniques. *J. Comput. Chem.*, 14:1212–1233, 1993.