Variational Integrators

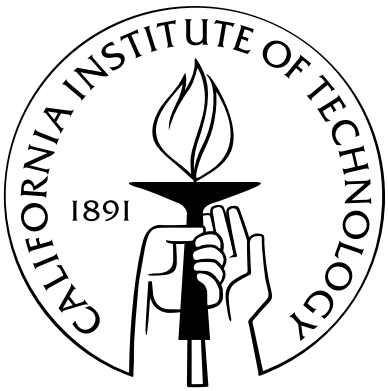
Thesis by

Matthew West

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2004

(Submitted May 28, 2004)

Acknowledgements

Many people have contributed to this work. In particular, I thank my advisor Jerry Marsden, with whom it has been a true pleasure and honour to work, and my co-authors Harish Bhat, Fehmi Cirak, Razvan Fetecau, Sameer Jalnapurkar, Couro Kane, Sanjay Lall, Melvin Leok, Adrian Lew, Marcel Oliver, Michael Ortiz, Sergey Pekarsky, Steve Shkoller, and Claudia Wulff, all of whom have contributed markedly to my education and enjoyment of research.

In addition, I would like to thank Darryl Holm, Arieh Iserles, Ben Leimkuhler, Christian Lubich, Robert McLachlan, Richard Murray, Sebastian Reich, Mark Roberts, Peter Schröder, Bob Skeel, and Yuri Suris for their valuable suggestions and help along the way.

Finally, this work would not have been possible without the love and support of my parents Owen and Judy, my sisters Kate and Anna, and of course Nicole.

Abstract

Variational integrators are a class of discretizations for mechanical systems which are derived by discretizing Hamilton's principle of stationary action. They are applicable to both ordinary and partial differential equations, and to both conservative and forced problems. In the absence of forcing they conserve (multi-)symplectic structures, momenta arising from symmetries, and energy up to a bounded error.

In this thesis the basic theory of discrete variational mechanics for ordinary differential equations is developed in depth, and is used as the basis for constructing variational integrators and analyzing their numerical properties. This is then taken as the starting point for the development of a new class of asynchronous time stepping methods for solid mechanics, known as Asynchronous Variational Integrators (AVIs). These explicit methods time step different elements in a finite element mesh with fully independent and decoupled time steps, allowing the simulation to proceed locally at the fastest rate allowed by local stability restrictions. Numerical examples of AVIs are provided, demonstrating the excellent properties they posess by virtue of their variational derivation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Variational integrators

This thesis futher develops the subject of variational integrators as it applies to mechanical systems of engineering interest. The idea behind this class of algorithms is to discretize the variational formulation of a given problem, rather than the differential equations. These problems may be either conservative or dissipative and forced, and may be cast as either ordinary or partial differential equations. For conservative problems, we focus on discretizing Hamilton's principle of stationary action in Lagrangian mechanics, while for dissipative or forced problems we discretize the Lagrange-d'Alembert principle. While the idea of discretizing variational formulations of mechanics is standard for elliptic problems, in the form of Galerkin and finite element methods (e.g., Johnson [1987], Hughes [1987]), it has only been applied relatively recently to derive variational time stepping algorithms for mechanical systems.

Advantages of variational integrators. The variational method for deriving integrators means that the resulting algorithms automatically have a number of properties. In particular, they are symplectic methods[1], they exactly preserve momenta associated to symmetries of the system, and they have excellent longtime energy stability.

These properties make them ideal for simulating physical systems which are either conservative or near-conservative. In such cases, one frequently wishes to compute certain averaged or statistical properties of the system, such as the speed of a shock-front, and maintaining good conservation properties where appropriate can be essential.

In addition, the variational methodology allows one to easily and cleanly derive good integrators even in extremely complex geometries, such as the asynchronous space-time meshes used in §6.1. This applies whether the mechanical system is conservative or not.

Finally, the variational methodology serves as a unifying framework from which to consider the

[1]Symplecticity is explained in more detail in §1.6.3, §1.6.4, §2.1.3, and §2.2.2.

Lagrangian side of symplectic integration theory. This fact derives from generating function theory, which shows that all symplectic integrators for Lagrangian systems are in fact variational.

Geometric integration. Variational integrators are examples of methods which preserve geometric structures of the continuous system, and as such they fall within the general class of *geometric integration* techniques. For mechanical systems the primary quantities on which most work has focussed are momenta, energy, and symplectic structures. An important result due to Ge and Marsden [1988] shows that an integrator can either preserve energy and momenta, or symplectic structure and momenta, but not all three. For this reason, the geometric integration community has largely split into those working on symplectic-momentum methods, including variational integrators, and those working on energy-momentum methods.

PDEs and multisymplectic discretizations. Variational integrators for ODEs can be extended to PDEs by discretizing both space and time variationally. An elegant framework within which this can be performed is given by *multisymplectic mechanics*, which is a local space-time viewpoint of mechanics. As with ODEs, variational methods for multisymplectic PDEs conserve (multi-) symplecticity, momenta, and have excellent energy behavior.

Asynchronous Variational Integrators (AVIs). In this thesis, the primary application of variational discretizations of multisymplectic PDEs is the development of Asynchronous Variational Integrators (AVIs) for solid mechanics. These methods use fully decoupled asynchronous space-time meshes, which means that in a finite element setting all elements can be time stepped independently of each other, subject only to local CFL conditions. This provides a large efficiency gain, as it is no longer necessary for the time step of the whole mesh to be dictated by the smallest element, as would ordinarily be the case for explicit time steppers. Furthermore, the variational method of derivation ensures that they asynchronicity is introduced without losing any of the good properties of standard methods such as the Newmark algorithm.

1.2 History and literature

Of course, the variational view of mechanics goes back to Euler, Lagrange and Hamilton. The form of the variational principle most important for continuous mechanics that we use in this article is due, of course, to Hamilton [1834]. We refer to Marsden and Ratiu [1999] for additional history, references and background on geometric mechanics.

There have been many attempts at the development of a discrete mechanics and corresponding integrators that we will not attempt to survey in any systematic fashion. The theory of discrete variational mechanics in the form we shall use it (that uses two copies $Q \times Q$ of the configuration space

for the discrete analogue of the velocity phase space) has its roots in the optimal control literature of the 1960s: see, for example, Jordan and Polak [1964], Hwang and Fan [1967] and Cadzow [1970]. In the context of mechanics early work was done, often independently, by Cadzow [1973], Logan [1973], Maeda [1980, 1981a,b], and Lee [1983, 1987], by which point the discrete action sum, the discrete Euler-Lagrange equations and the discrete Noether's theorem were clearly understood. This theory was then pursued further in the context of integrable systems in Veselov [1988, 1991] and Moser and Veselov [1991], and in the context of quantum mechanics in Jaroszkiewicz and Norton [1997a,b] and Norton and Jaroszkiewicz [1998].

The variational view of discrete mechanics and its numerical implementation is further developed in Wendlandt and Marsden [1997a,b] and then extended in Kane, Marsden, and Ortiz [1999a], Marsden, Pekarsky, and Shkoller [1999b,a], Bobenko and Suris [1999a,b] and Kane, Marsden, Ortiz, and West [2000]. The beginnings of an extension of these ideas to a nonsmooth framework is given in Kane, Repetto, Ortiz, and Marsden [1999b], and is carried further in Fetecau, Marsden, Ortiz, and West [2003a]. Other applications include Rowley and Marsden [2002], and an investigation of convergence is given in Müller and Ortiz [2003].

Other discretizations of Hamilton's principle are given in Mutze [1998], Cano and Lewis [1998] and Shibberu [1994]. Other versions of discrete mechanics (not necessarily discrete Hamilton's principles) are given in, for instance, Itoh and Abe [1988], Labudde and Greenspan [1974, 1976a,b], and MacKay [1992].

Of course, there have been many works on symplectic integration, largely done from other points of view than that developed here. We will not attempt to survey this in any systematic fashion, as the literature is simply too large with too many points of view and too many intricate subtleties. We give a few highlights and give further references in the body of the thesis. For instance, we shall connect the variational view with the generating function point of view that was begun in De Vogelaére [1956]. Generating function methods were developed and used in, for example, Ruth [1983], Forest and Ruth [1990] and in Channell and Scovel [1990]. See also Berg, Warnock, Ruth, and Forest [1994], and Warnock and Ruth [1992, 1991]. For an overview of symplectic integration, see Hairer, Lubich, and Wanner [2002], as well as Sanz-Serna [1992b] and Sanz-Serna and Calvo [1994]. Qualitative properties of symplectic integration of Hamiltonian systems are given in Gonzalez, Higham, and Stuart [1999] and Cano and Sanz-Serna [1997]. Longtime energy behaviour for oscillatory systems is studied in Hairer and Lubich [2000]. Longtime behaviour of symplectic methods for systems with dissipation is given in Hairer and Lubich [1999]. A numerical study of preservation of adiabatic invariants is given in Reich [1999b] and Shimada and Yoshida [1996]. Backward error analysis is studied in Benettin and Giorgilli [1994], Hairer [1994], Hairer and Lubich [1997] and Reich [1999a]. Other ideas connected to the above literature include those of Baez and Gilliam [1994], Gilliam [1996], Gillilan and Wilson [1992]. For other references see the large literature on symplectic methods in

molecular dynamics, such as Schlick, Skeel, Brunger, Kale, Board, Hermans, and Schulten [1999], and for various applications, see Hardy, Okunbor, and Skeel [1999], Leimkuhler and Skeel [1994], Barth and Leimkuhler [1996] and references therein.

A single-step variational idea that is relevant to our approach is given in Ortiz and Stainier [1999], and developed further in Radovitzky and Ortiz [1999], and Kane et al. [1999b, 2000].

Direct discretizations on the Hamiltonian side, where one discretizes the Hamiltonian and the symplectic structure, are developed in Gonzalez [1996b,a] and further in Gonzalez [1999] and Gonzalez et al. [1999]. This is developed and generalized much further in McLachlan, Quispel, and Robidoux [1998, 1999].

Multisymplectic mechanics has its origins in the mid-twentieth century (see Kijowski and Tulczyjew [1979] and references therein for a representative view). There has been a recent explosion in work, however, driven by numerical applications, beginning with independent work by Marsden, Patrick, and Shkoller [1998] and Bridges and Reich [2001a]. Non-numerical applications have also advanced, with work by Marsden and Shkoller [1999], Kouranbaeva and Shkoller [2000], Bridges and Laine-Pearson [2001], Hydon [2001], Bridges and Derks [2002], and Binz, de Leon, de Diego, and Socolescu [2002]. The work on numerical applications has advanced on several main fronts, including that in Bridges and Reich [2001b], Reich [2000a], and Reich [2000b], as well as Islas and Schober [2002] and Islas, Karpeev, and Schober [2001], and the series of papers by Sun and Qin [2000], Wang and Qin [2001], Guo, Li, and Wu [2001b], Chen [2001], Guo, Ji, Li, and Wu [2001a], Guo, Li, Wu, and Wang [2002a], Guo, Li, Wu, and Wang [2002b], Guo, Li, Wu, and Wang [2002c], Chen and Qin [2002], Liu and Qin [2002], Chen [2002], Hong and Qin [2002], Wang and Qin [2002], and Chen [2003]. Finally, there is the work associated with this thesis, which is discussed in the next section.

The asynchronous methods developed in this thesis have much in common with multi-time step integration algorithms, sometimes termed subcycling methods. These algorithms have been developed in Neal and Belytschko [1989] and Belytschko and Mullen [1976], mainly to allow high-frequency elements to advance at smaller time steps than the low-frequency ones. In its original version, the method grouped the nodes of the mesh and assigned to each group a different time step. Adjacent groups of nodes were constrained to have integer time step ratios (see Belytschko and Mullen [1976]), a condition that was relaxed in Neal and Belytschko [1989] and Belytschko [1981]. Recently an implicit multi-time step integration method was developed and analyzed in Smolinski and Wu [1998]. We also mention the related work of Hughes and Liu [1978] and Hughes, Pister, and Taylor [1979]. There are also many connections between the multi-time step impulse method (also known as Verlet-I and r-RESPA), which is popular in molecular dynamics applications, and the AVI algorithm (see Grubmüller, Heller, Windemuth, and Schulten [1991] and Tuckerman, Berne, and Martyna [1992]).

1.3 Work associated with this thesis

Work incorporated in and arising from this thesis is:

- Kane, Marsden, Ortiz, and West [2000] studied the Newmark algorithm from the perspective of variational integrators and proved that the Newmark method with $\gamma = 1/2$ is itself variational and hence symplectic. This paper also included the first introduction of the idea of a discrete Lagrange-d'Alembert principle for forced and dissipative systems.

- Marsden, Pekarsky, Shkoller, and West [2001] laid the groundwork for variational multisymplectic integrators of continuum theories by reformulating classical continuum mechanics in an intrinsic multisymplectic framework. The two examples treated in detail were ideal fluids and hyper-elasticity, and much attention was paid to the role of incompressibility. This necessitated the development of the first theory of constrained multisymplectic systems.

- Marsden and West [2001] provided a survey of the existing theory for discrete variational mechanics and variational integrators for ODEs. It also presented a great deal of new theory for the first time, including much work on the numerical properties of variational integrators such as high-order methods and approximation accuracy results.

- Fetecau, Marsden, Ortiz, and West [2003a] considered Lagrangian systems with trajectories which are continuous but nonsmooth, with collision problems being the main focus. This lead to the development of both a continuous variational theory for such systems as well as variational integrators for such problems. These were the first geometric integrators developed which were applicable to nonsmooth systems.

- Fetecau, Marsden, and West [2003b] extended the nonsmooth ODE theory from Fetecau et al. [2003a] to the multisymplectic PDE setting. This provides a formulation which encompasses such problems as fluid-solid interactions, internal shock waves in continua and collisions of elastic bodies.

- Lew, Marsden, Ortiz, and West [2003a] introduced the concept of Asynchronous Variational Integrators (AVI). This paper also developed further the concept of full variations for multisymplectic systems and the conservation properties associated with horizontal variations.

- Lew, Marsden, Ortiz, and West [2003b] further investigated the numerics of AVIs. This involved studying a large scale simulation of a helicopter rotor blade as well as providing a proof of convergence and an analysis of the computational complexity of AVIs. This was the first proof of convergence for a fully asynchronous integration time integration method.

- Jalnapurkar, Leok, Marsden, and West [2003] developed a discrete version of Routh reduction theory for systems with abelian symmetry groups.

- Oliver, West, and Wulff [2003] studied the approximate momentum conservation properties of variational discretizations of nonlinear wave equations, giving both numerical examples and theory explaining the observed behavior. This represented the first rigorous results concerning the backward error analysis of spatial PDE discretizations.

- Cirak and West [2003] used the ideas in Fetecau et al. [2003a] as the basis for constructing a very efficient technique for explicit time stepping of finite element models with collisions, and demonstrated the algorithms on large parallel simulations of colliding shells and solids.

- Pekarsky and West [2003] developed groupoid discretizations of diffeomorphism groups and used these to construct the first exactly circulation preserving integrators for ideal fluids.

- Lall and West [2003] unified the discretizations of the calculus of variations used in discrete mechanics and in discrete optimal control. This lead to the development of the Hamiltonian side of discrete mechanics and the concept of a discrete Hamilton-Jacobi equation for discrete mechanics.

1.4 Outline of thesis

We begin in §1.5 to §1.7 with a simple overview of variational integrators for ODEs and PDEs. This material is a summary of §2 to §6.

In §2 we develop discrete variational mechanics for ODEs, including extensive comparisons with continuous-time Lagrangian and Hamiltonian mechanics. This is then used in §3 as the basis for variational integrators, whose numerical properties are investigated in detail. Next, §4 considers both discrete mechanics and variational integrators for systems with forcing and constraints.

The final two chapters deal with variational mechanics and integrators for PDEs. In §5 we formulate continuum mechanics within the context of variational multisymplectic mechanics, while in §6 we develop Asynchronous Variational Integrators (AVIs) as a special case of variational multisymplectic discretizations.

1.5 Discrete Dynamics and Variational Integrators

In this section we give a brief overview of how discrete variational mechanics can be used to derive variational integrators. We begin by reviewing the derivation of the Euler-Lagrange equations, and then show how to mimic this process on a discrete level.

1.5.1 Continuous time Lagrangian dynamics

For concreteness, consider the Lagrangian system $L(q,\dot{q}) = \frac{1}{2}\dot{q}^{\mathsf{T}} M\dot{q} - V(q)$, where M is a symmetric positive-definite mass matrix and V is a potential function. We work in $\mathbb{R}^n$ or in generalized coordinates and will use vector notation for simplicity, so $q = (q^1, q^2, \ldots, q^n)$. In the standard approach of Lagrangian mechanics, we form the action function by integrating L along a curve $q(t)$ and then compute variations of the action while holding the endpoints of the curve $q(t)$ fixed. This gives

$$\begin{aligned}\delta S(q) = \delta \int_0^T L\big(q(t),\dot{q}(t)\big)\,dt &= \int_0^T \left[\frac{\partial L}{\partial q}\cdot\delta q + \frac{\partial L}{\partial \dot{q}}\cdot\delta\dot{q}\right] dt \\ &= \int_0^T \left[\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right)\right]\cdot\delta q\;dt + \left[\frac{\partial L}{\partial \dot{q}}\cdot\delta q\right]_0^T, \end{aligned} \tag{1.1}$$

where we have used integration by parts. The final term is zero because we assume that $\delta q(T) = \delta q(0) = 0$. Requiring that the variations of the action be zero for all δq implies that the integrand must be zero for each time t, giving the well-known ***Euler-Lagrange equations***

$$\frac{\partial L}{\partial q}(q,\dot{q}) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}(q,\dot{q})\right) = 0. \tag{1.2}$$

For the particular form of the Lagrangian chosen above, this is just

$$M\ddot{q} = -\nabla V(q),$$

which is Newton's equation: mass times acceleration equals force. It is well known that the system described by the Euler-Lagrange equations has many special properties. In particular, the flow on state space is symplectic, meaning that it conserves a particular two-form, and if there are symmetry actions on phase space, then there are corresponding conserved quantities of the flow, known as momentum maps. We will return to these ideas later in this work in §2.

1.5.2 Discrete time Lagrangian dynamics

We will now see how discrete variational mechanics performs an analogue of the above derivation. Rather than taking a position q and velocity $\dot{q}$, consider now two positions q_0 and q_1 and a time step $\Delta t \in \mathbb{R}$. These positions should be thought of as being two points on a curve at time Δt apart, so that $q_0 \approx q(0)$ and $q_1 \approx q(\Delta t)$.

We now consider a discrete Lagrangian $L_d(q_0, q_1, \Delta t)$, which we think of as approximating the action integral along the curve segment between q_0 and q_1. For concreteness, consider the very simple approximation to the integral $\int_0^T L\,dt$ given by using the rectangle rule[2] (the length of the

[2] As we shall see later, more sophisticated quadrature rules lead to higher-order accurate integrators.

interval times the value of the integrand with the velocity vector replaced by $(q_1 - q_0)/\Delta t$):

$$L_d(q_0, q_1, \Delta t) = \Delta t \left[\frac{1}{2} \left(\frac{q_1 - q_0}{\Delta t} \right)^{\mathsf{T}} M \left(\frac{q_1 - q_0}{\Delta t} \right) - V(q_0) \right]. \tag{1.3}$$

Next consider a discrete curve of points $\{q_k\}_{k=0}^N$ and calculate the discrete action along this sequence by summing the discrete Lagrangian on each adjacent pair. Following the continuous derivation above, we compute variations of this action sum with the boundary points q_0 and q_N held fixed. This gives

$$\begin{aligned} \delta S_d(\{q_k\}) = \delta \sum_{k=0}^{N-1} L_d(q_k, q_{k+1}, \Delta t) &= \sum_{k=0}^{N-1} \left[D_1 L_d(q_k, q_{k+1}, \Delta t) \cdot \delta q_k + D_2 L_d(q_k, q_{k+1}, \Delta t) \cdot \delta q_{k+1} \right] \\ &= \sum_{k=1}^{N-1} \left[D_2 L_d(q_{k-1}, q_k, \Delta t) + D_1 L_d(q_k, q_{k+1}, \Delta t) \right] \cdot \delta q_k \\ &\quad + D_1 L_d(q_0, q_1, \Delta t) \cdot \delta q_0 + D_2 L_d(q_{N-1}, q_N, \Delta t) \cdot \delta q_N, \end{aligned} \tag{1.4}$$

where we have used a discrete integration by parts (rearranging the summation). Henceforth, $D_i L_d$ indicates the slot derivative with respect ot the i-th argument of L_d. If we now require that the variations of the action be zero for any choice of δq_k with $\delta q_0 = \delta q_N = 0$, then we obtain the ***discrete Euler-Lagrange equations***

$$D_2 L_d(q_{k-1}, q_k, \Delta t) + D_1 L_d(q_k, q_{k+1}, \Delta t) = 0, \tag{1.5}$$

which must hold for each k. For the particular L_d chosen above, we compute

$$\begin{aligned} D_2 L_d(q_{k-1}, q_k, \Delta t) &= M \left(\frac{q_k - q_{k-1}}{\Delta t} \right) \\ D_1 L_d(q_k, q_{k+1}, \Delta t) &= - \left[M \left(\frac{q_{k+1} - q_k}{\Delta t} \right) + (\Delta t) \nabla V(q_k) \right], \end{aligned}$$

and so the discrete Euler-Lagrange equations are

$$M \left(\frac{q_{k+1} - 2q_k + q_{k-1}}{(\Delta t)^2} \right) = -\nabla V(q_k).$$

This is clearly a discretization of Newton's equations, using a simple finite difference rule for the derivative.

If we take initial conditions (q_0, q_1), then the discrete Euler-Lagrange equations define a recursive rule for calculating the sequence $\{q_k\}_{k=0}^N$. Regarded in this way, they define a map $(q_k, q_{k+1}) \mapsto (q_{k+1}, q_{k+2})$ which we can think of as a one-step integrator for the system defined by the continuous Euler-Lagrange equations. This viewpoint is considered in depth in §3.

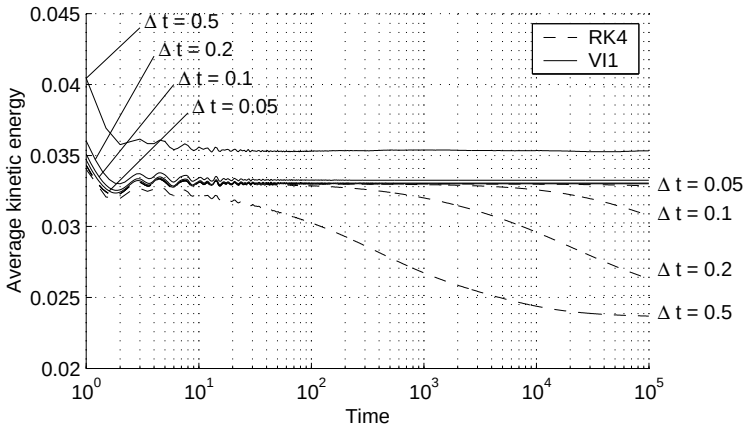


Figure 1.1: Average kinetic energy (1.6) as a function of T for a nonlinear spring-mass lattice system, using a first-order variational integrator (VI1) and a fourth-order Runge-Kutta method (RK4) and a range of time steps Δt. Observe that the Runge-Kutta method suffers substantial numerical dissipation, unlike the variational method.

Heat calculation example. As we will consider in detail in §1.6, variational integrators are interesting because they inherit many of the conservative properties of the original Lagrangian system. As an example of this, we consider the numerical approximation of the heat of a coupled spring-mass lattice model. The numerical heat for time T is defined to be the numerical approximation of

$$\bar{K}(T) = \frac{1}{T}\int_0^T \frac{1}{2}\|\dot{q}\|^2\,dt, \tag{1.6}$$

while the true heat of the system is the limit of the quantity,

$$\bar{K} = \lim_{T\to\infty} \bar{K}(T). \tag{1.7}$$

The temperature of the system, which is an intensive—as opposed to extensive—quantity, is the heat K divided by the heat capacity nd, where n is the number of masses and d is the dimension of space. We assume that the system is ergodic and that this limit exists. In Figure 1.1 we plot the numerical approximations to (1.6) at $T = 10^5$ computed using a first-order variational integrator (VI1) and a fourth-order Runge-Kutta method (RK4). As the time step is decreased the numerical solution tends towards the true solution.

Note, however, that the lack of dissipation in the variational integrator means that for quite large time steps it computes the averaged kinetic energy much better. To make this precise, we consider the harmonic approximation to the lattice system (that is, the linearization), for which we can compute the limit (1.7) analytically. The error in the numerically computed heat is plotted in Figure 1.2 for a range of different time steps Δt and final times T, using the same first-order variational method (VI1) and fourth-order Runge-Kutta method (RK4), as well as a fourth-order

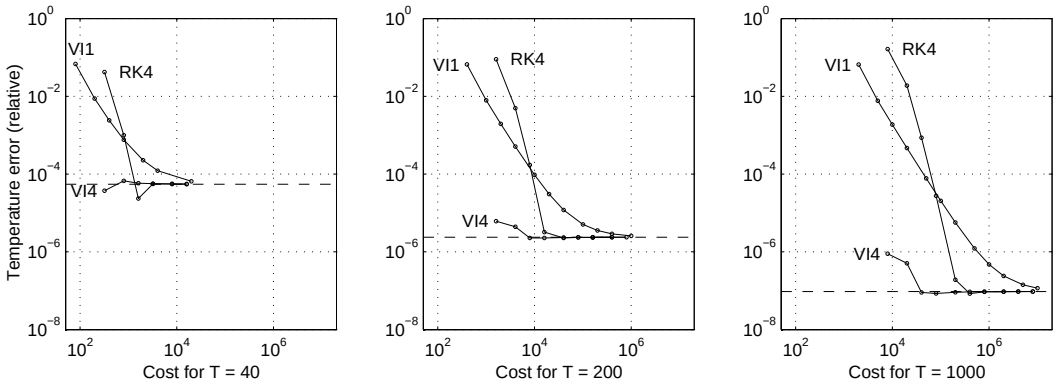


Figure 1.2: Error in numerically approximated heat for the harmonic (linear) approximation to the lattice system from Figure 1.1, using a first-order variational integrator (VI1), a fourth-order Runge-Kutta method (RK4) and a fourth-order variational integrator (VI4). The three plots have different final times T, while the cost is increased within each plot by decreasing the time step Δt. For each T the dashed horizontal line is the exact value of $\bar{K}(T) - \bar{K}$, which is the minimum error that the numerical approximation can achieve without increasing T. Observe that the low-order variational method VI1 beats the traditional RK4 method for larger errors, while the high-order variational method VI4 combines the advantages of both high-order and variational structure to always win.

variational integrator (VI4).

To compute the heat (1.7) numerically we must clearly let $T \to \infty$ and $\Delta t \to 0$. Both of these limits increase the cost of the simulation, and so there is a tradeoff between them. As we see in Figure 1.2, for a fixed T there is some Δt which adequately resolves the integral (1.6), and so the error cannot decrease any further without increasing T. To see this, take a numerical approximation $\bar{K}(T, \Delta t)$ to (1.6) and decompose the error as

$$\underbrace{\bar{K}(T,\Delta t) - \bar{K}}_{\text{total error}} = \underbrace{\bar{K}(T,\Delta t) - \bar{K}(T)}_{\text{discretization error}} + \underbrace{\bar{K}(T) - \bar{K}}_{\text{limit error}}. \tag{1.8}$$

Decreasing Δt will reduce the discretization error, but at some point this will become negligible compared to the limit error, which will only tend to zero as T is increased.

The striking feature of Figure 1.2 is that the variational integrators perform far better than a traditional Runge-Kutta method. For large error tolerances, such as 1% or 5% error (10^{-2} or 5×10^{-2} in Figure 1.2), the first-order variational method is very cheap and simple. For higher precision, the fourth-order Runge-Kutta method eventually becomes cheaper than the first-order variational integrator, but the fourth-order variational method combines the advantages of both and is always the method of choice.

Of course, such sweeping statements as above have to be interpreted and used with great care, as in the precise statements in the text that follows. For example, if the integration step size is too large, then sometimes energy can behave very badly, even for a variational integrator (see, for

example, Gonzalez and Simo [1996]). It is likewise well known that energy conservation does not guarantee trajectory accuracy. These points will be discussed further below.

1.5.3 Variational integrators

We are primarily interested in discrete Lagrangian mechanics for deriving integrators for mechanical systems. Any integrator which is the discrete Euler-Lagrange equation for some discrete Lagrangian is called a ***variational integrator***. As we have seen above, variational integrators can be implemented by taking two configurations q_0 and q_1 of the system, which should approximate $q_0 \approx q(t_0)$ and $q_1 \approx q(t_0 + \Delta t)$, and then solving the discrete Euler-Lagrange equations (1.5) for q_2. This process can then be repeated to calculate an entire discrete trajectory. The map $(q_{k-1}, q_k) \mapsto (q_k, q_{k+1})$ defined by the discrete Euler-Lagrange equations is known as the ***discrete evolution map***.

Position-momentum form. For mechanical systems it is more common to specifiy the initial conditions as a position and a velocity (or momentum), rather than two positions. To rewrite a variational integrator in a position-momentum form we first observe that we can define the momentum at time step k to be

$$p_k = D_2 L_d(q_k, q_{k+1}, \Delta t) = -D_1 L_d(q_{k-1}, q_k, \Delta t). \tag{1.9}$$

The two expressions for p_k are equal because this equality is precisely the discrete Euler-Lagrange equations (1.5). Using this definition we can write the ***position-momentum form*** of a variational integrator as

$$p_k = -D_1 L_d(q_k, q_{k+1}, \Delta t) \tag{1.10a}$$

$$p_{k+1} = D_2 L_d(q_k, q_{k+1}, \Delta t). \tag{1.10b}$$

Given an initial condition (q_0, p_0) we can solve the implicit equation (1.10a) to find q_1, and then evaluate (1.10b) to give p_1. We then have (q_1, p_1) and we can repeat the procedure. The sequence $\{q_k\}_{k=0}^N$ so obtained will clearly satisfy the regular discrete Euler-Lagrange equations (1.5) for all k, due to the definition (1.9) of p_k. This equality is further elaborated in §2.4.2.

Order of accuracy. We remarked above that a discrete Lagrangian should be thought of as approximating the continuous action integral. We will now make this statement precise. We say that a discrete Lagrangian is of ***order*** r if

$$L_d(q_0, q_1, \Delta t) = \int_0^{\Delta t} L(q(t), \dot{q}(t)) dt + \mathcal{O}(\Delta t)^{r+1}, \tag{1.11}$$

where $q(t)$ is the unique solution of the Euler-Lagrange equations for L with $q(0) = q_0$ and $q(\Delta t) = q_1$. It can then be proven (see Theorem 3.3) that if L_d is of order r, then the corresponding variational integrator is also of order r, so that

$$q_k = q(k\,\Delta t) + \mathcal{O}(\Delta t)^{r+1}.$$

To design high-order variational integrators, we must therefore construct discrete Lagrangians which accurately approximate the action integral.

Symmetric methods. One useful observation when calculating the order of integrators is that symmetric methods always have even order. We say that a discrete Lagrangian is ***symmetric*** if

$$L_d(q_0, q_1, \Delta t) = -L_d(q_1, q_0, -\Delta t). \tag{1.12}$$

This implies [Marsden and West, 2001, Theorem 2.4.1] that the resulting variational integrator will also be symmetric, and will thus automatically be of even order. We will use this fact below.

Geometric aside. The definition (1.9) of p_k defines a map $Q \times Q \to T^*Q$. In fact we can define two such maps, known as the discrete Legendre transforms, by $\mathbb{F}L_d^+(q_0, q_1) = (q_1, D_2 L_d(q_0, q_1, \Delta t))$ and $\mathbb{F}L_d^-(q_0, q_1, \Delta t) = (q_1, -D_1 L_d(q_0, q_1))$. This is discussed further in §2.4.1. The position-momentum form (1.10) of the discrete Euler-Lagrange equations is thus given by $\tilde{F}_{L_d}^{\Delta t} = \mathbb{F}L_d^{\pm} \circ F_{L_d}^{\Delta t} \circ (\mathbb{F}L_d^{\pm})^{-1}$ and is a map $\tilde{F}_{L_d}^{\Delta t} : T^*Q \to T^*Q$, where $F_{L_d}^{\Delta t} : Q \times Q \to Q \times Q$ is the discrete evolution map. This shows that variational integrators are really one-step methods, although they may initially appear to be two-step. This form of the integrator is called the discrete Hamilton map and is investigated in §2.4.3.

1.5.4 Examples of discrete Lagrangians

We now consider some examples of discrete Lagrangians.

Generalized midpoint rule. The classical midpoint rule for the system $\dot{x} = f(x)$ is given by $x_{k+1} - x_k = (\Delta t) f((x_{k+1} + x_k)/2)$. If we add a parameter $\alpha \in [0, 1]$ where the force evaluation occurs (so $\alpha = 1/2$ is the standard midpoint), then we can write the corresponding discrete Lagrangian

$$\begin{aligned} L_d^{\mathrm{mp},\alpha}(q_0, q_1, \Delta t) &= (\Delta t) L\left((1-\alpha)q_0 + \alpha q_1, \frac{q_1 - q_0}{\Delta t}\right) \\ &= \frac{\Delta t}{2}\left(\frac{q_1 - q_0}{\Delta t}\right)^{\mathsf{T}} M \left(\frac{q_1 - q_0}{\Delta t}\right) - (\Delta t) V\Big((1-\alpha)q_0 + \alpha q_1\Big). \end{aligned} \tag{1.13}$$

The discrete Euler-Lagrange equations (1.5) are thus

$$M\left(\frac{q_{k+1}-2q_k+q_{k-1}}{(\Delta t)^2}\right) = -(1-\alpha)\nabla V\Big((1-\alpha)q_k+\alpha q_{k+1}\Big) - \alpha\nabla V\Big((1-\alpha)q_{k-1}+\alpha q_k\Big) \quad (1.14)$$

and the position-momentum form (1.10) of the variational integrator is

$$p_k = M\left(\frac{q_{k+1}-q_k}{\Delta t}\right) + (1-\alpha)(\Delta t)\nabla V\Big((1-\alpha)q_k+\alpha q_{k+1}\Big) \quad (1.15a)$$

$$p_{k+1} = M\left(\frac{q_{k+1}-q_k}{\Delta t}\right) - \alpha(\Delta t)\nabla V\Big((1-\alpha)q_k+\alpha q_{k+1}\Big). \quad (1.15b)$$

This is always an implicit method, and for general $\alpha \in [0,1]$ it is first-order accurate. When $\alpha = 1/2$ it is easy to see that $L_d^{\mathrm{mp},\alpha}$ is symmetric, and thus the integrator is second-order.

Generalized trapezoidal rule. Rather than evaluating the force at an averaged location, we could instead average the evaluated forces. Doing so at a parameter $\alpha \in [0,1]$ gives a generalization of the trapezoidal rule

$$\begin{aligned} L_d^{\mathrm{tr},\alpha}(q_0,q_1,\Delta t) &= (\Delta t)(1-\alpha)L\left(q_0,\frac{q_1-q_0}{\Delta t}\right) + (\Delta t)\alpha L\left(q_1,\frac{q_1-q_0}{\Delta t}\right) \\ &= \frac{\Delta t}{2}\left(\frac{q_1-q_0}{\Delta t}\right)^{\mathsf{T}} M\left(\frac{q_1-q_0}{\Delta t}\right) - (\Delta t)\Big((1-\alpha)V(q_0)+\alpha V(q_1)\Big). \end{aligned} \quad (1.16)$$

Computing the discrete Euler-Lagrange equations (1.5) gives

$$M\left(\frac{q_{k+1}-2q_k+q_{k-1}}{(\Delta t)^2}\right) = -\nabla V(q_k) \quad (1.17)$$

with corresponding position-momentum (1.10) form

$$p_k = M\left(\frac{q_{k+1}-q_k}{\Delta t}\right) + (\Delta t)(1-\alpha)\nabla V(q_k) \quad (1.18a)$$

$$p_{k+1} = \left(\frac{q_{k+1}-q_k}{\Delta t}\right) - (\Delta t)\alpha\nabla V(q_k). \quad (1.18b)$$

This method is explicit for all α, and is generally first-order accurate. For $\alpha = 1/2$ it is symmetric, and thus becomes second-order accurate.

Observe that there is no α in the discrete Euler-Lagrange equations (1.17), although it does appear in the position-momentum form (1.18). This means that the only effect of α is on the starting procedure of this integrator, as thereafter the trajectory will be entirely determined by (1.17). If we are given an initial position and momentum (q_0,p_0), then we can use (1.18a) to calculate q_1 and then continue with (1.17) for future time steps. For this procedure to be second-order accurate it is necessary to take $\alpha = 1/2$ in the use of (1.18a) for the first time step.

Newmark method. The Newmark family of integrators, originally given in Newmark [1959], are widely used in structural dynamics codes. They are usually written (see, for example, Hughes [1987]) for the system $L = \frac{1}{2}\dot{q}^{\mathsf{T}} M \dot{q} - V(q)$ as maps $(q_k, \dot{q}_k) \mapsto (q_{k+1}, \dot{q}_{k+1})$ satisfying the implicit relations

$$q_{k+1} = q_k + (\Delta t)\dot{q}_k + \frac{1}{2}(\Delta t)^2 \left[(1-2\beta)a(q_k) + 2\beta a(q_{k+1})\right] \tag{1.19a}$$

$$\dot{q}_{k+1} = \dot{q}_k + (\Delta t)\left[(1-\gamma)a(q_k) + \gamma a(q_{k+1})\right] \tag{1.19b}$$

$$a(q) = M^{-1}(-\nabla V(q)), \tag{1.19c}$$

where the parameters $\gamma \in [0,1]$ and $\beta \in [0, \frac{1}{2}]$ specify the method. It is simple to check that the method is second-order if $\gamma = 1/2$ and first-order otherwise, and that it is generally explicit only for $\beta = 0$.

The $\beta = 0$, $\gamma = 1/2$ case is well known to be symplectic (see, for example, Simo, Tarnow, and Wong [1992]) with respect to the canonical symplectic form Ω_L. This can be easily seen from the fact that this method is a rearrangement of the position-momentum form of the generalized trapezoidal rule with $\alpha = 1/2$. Note that this method is the same as the velocity Verlet method, which is popular in molecular dynamics codes. As we remarked above, if the method (1.18) is implemented by taking one initial step with (1.18a) as a starting procedure, and then continued with (1.17), then this will give a method essentially equivalent to explicit Newmark. To be exactly equivalent, however, and to be second-order accurate, one must take $\alpha = 1/2$ in the use of (1.18a). This will be of importance in §6.1.6.

It is also well known (for example, Simo et al. [1992]) that the Newmark algorithm with $\beta \neq 0$ does not preserve the *canonical* symplectic form. Nonetheless it can be shown Kane et al. [2000] that the Newmark method with $\gamma = 1/2$ and any β can be generated from a discrete Lagrangian, and it thus preserves a non-canonical symplectic structure. An alternative and independent method of analyzing the symplectic members of Newmark has been given by Skeel, Zhang, and Schlick [1997], including an interesting nonlinear analysis in Skeel and Srinivas [2000]. The Newmark method is discussed in greater detail in §3.6.3.

Galerkin methods and symplectic Runge-Kutta schemes. Both the generalized midpoint and generalized trapezoidal discrete Lagrangians discussed above can be viewed as particular cases of linear finite element discrete Lagrangians. If we take shape functions

$$\phi_0(\alpha) = 1 - \alpha \qquad \phi_1(\alpha) = \alpha, \tag{1.20}$$

then a general linear Galerkin discrete Lagrangian is given by

$$L_d^{G,0}(q_0, q_1, \Delta t) = \sum_{i=1}^{m} w_i L\left(\phi_0(\alpha_i)q_0 + \phi_1(\alpha_i)q_1, \frac{\dot{\phi}_0(\alpha_i)q_0 + \dot{\phi}_1(\alpha_i)q_1}{\Delta t}\right), \tag{1.21}$$

where (α_i, w_i), $i = 1, \ldots, m$, is a set of quadrature points and weights. Taking $m = 1$ and $(\alpha_1, w_1) = (\alpha, 1)$ gives the generalized midpoint rule, while taking $m = 2$, $(\alpha_1, w_1) = (0, 1-\alpha)$ and $(\alpha_2, w_2) = (1, \alpha)$ gives the generalized trapezoidal rule.

Taking high-order finite element basis functions and quadrature rules is one method to construct high-order variational integrators. In general, we have a set of basis functions ϕ_j, $j = 0, \ldots, s$, and a set of quadrature points (α_i, w_i), $i = 1, \ldots, m$. The resulting Galerkin discrete Lagrangian is then

$$L_d^{G,s,\text{full}}(q_0, \ldots, q_s, \Delta t) = \sum_{i=1}^{m} w_i L\left(\sum_{j=0}^{s} \phi_j(\alpha_i)q_j, \frac{1}{\Delta t}\sum_{j=0}^{s} \dot{\phi}_j(\alpha_i)q_j\right). \tag{1.22}$$

This $(s+1)$-point discrete Lagrangian can be used to derive a standard two-point discrete Lagrangian by taking

$$L_d^{G,s}(q_0, q_1, \Delta t) = \operatorname*{ext}_{Q_1,\ldots,Q_{s-1}} L_d^{G,s,\text{full}}(q_0, Q_1, \ldots, Q_{s-1}, q_1, \Delta t), \tag{1.23}$$

where ext $L_d^{G,s,\text{full}}$ means that $L_d^{G,s,\text{full}}$ should be evaluated at extreme or critical values of $Q_1, \ldots, Q_s$. When $s = 1$ we immediately recover (1.21). Of course, using the discrete Lagrangian (1.22) is equivalent to a finite element discretization in time of (1.1), as in Bottasso [1997] for example.

An interesting feature of Galerkin discrete Lagrangians is that the resulting variational integrator can always be implemented as a partitioned Runge-Kutta method (see §3.6.6 for details). Using this technique high-order implicit methods can be constructed, including the collocation Gauss-Legendre family and the Lobatto IIIA-IIIB family of integrators.

1.5.5 Constrained systems

Many physical systems can be most easily expressed by taking a larger system and imposing constraints, which we take here to mean requiring that a given constraint function g is zero for all configurations. To discretize such problems, we can either work in local coordinates on the constraint set $\{q \mid g(q) = 0\}$, or we can work with the full configurations q and use Lagrange multipliers to enforce $g(q) = 0$. Here we consider the second option, as the first option requires no modification to the variational integrator theory[3].

[3] In the event that the constraint set is not a vector space, local coordinates would require the more general theory of discrete mechanics on smooth manifolds, as in §4.

Taking variations of the action with Lagrange multipliers added requires that

$$\delta \sum_{k=0}^{N-1} \Big[L_d(q_k, q_{k+1}, h) + \lambda_{k+1} \cdot g(q_{k+1}) \Big] = 0 \tag{1.24}$$

and so using (1.4) gives the ***constrained discrete Euler-Lagrange equations***

$$D_2 L_d(q_{k-1}, q_k, h) + D_1 L_d(q_k, q_{k+1}, h) = -\lambda_k \cdot \nabla g(q_k) \tag{1.25a}$$

$$g(q_{k+1}) = 0 \tag{1.25b}$$

which can be solved for λ_k and q_{k+1}[4]. These equations have all of the conservation properties, such as symplecticity and momentum conservation, as the unconstrained discrete equations.

An interesting example of a constrained variational integrator is the SHAKE method [Ryckaert, Ciccotti, and Berendsen, 1977], which can be neatly obtained by taking the generalized trapezoidal rule of §1.5.4 with $\alpha = 1/2$ and forming the constrained equations as in (1.25). This is carried out in §4.5.4.

1.5.6 Forcing and dissipation

Now we consider nonconservative systems; those with forcing and those with dissipation. For problems in which the nonconservative forcing dominates there is likely to be little benefit from variational integration techniques. There are many problems, however, for which the system is primarily conservative, but where there are very weak nonconservative effects which must be accurately accounted for. Examples include weakly damped systems, such as photonic drag on satellites, and small control forces, such as arise in continuous thrust technologies for spacecraft. In applications such as these the conservative behavior of variational integrators can be very important, as they do not introduce numerical dissipation in the conservative part of the system, and thus accurately resolve the small nonconservative forces.

Recall that the (continuous) ***integral Lagrange-d'Alembert principle*** is

$$\delta \int L(q(t), \dot{q}(t)) dt + \int F(q(t), \dot{q}(t)) \cdot \delta q \; dt = 0, \tag{1.26}$$

where $F(q, v)$ is an arbitrary force function. We define the ***discrete Lagrange-d'Alembert principle*** to be

$$\delta \sum L_d(q_k, q_{k+1}) + \sum \left[F_d^-(q_k, q_{k+1}) \cdot \delta q_k + F_d^+(q_k, q_{k+1}) \cdot \delta q_{k+1} \right] = 0, \tag{1.27}$$

[4]Observe that the linearization of the above system is not symmetric, unlike for constrained elliptic problems. This is because we are solving forward in time, rather than for all times at once as in a boundary value problem.

where L_d is the discrete Lagrangian and F_d^- and F_d^+ are the left and right discrete forces. These should approximate the continuous forcing so that

$$F_d^-(q_k, q_{k+1}) \cdot \delta q_k + F_d^+(q_k, q_{k+1}) \cdot \delta q_{k+1} \approx \int_{t_k}^{t_{k+1}} F(q(t), \dot{q}(t)) \cdot \delta q \; dt.$$

The equation (1.27) defines an integrator $(q_k, q_{k+1}) \mapsto (q_{k+1}, q_{k+2})$ given implicitly by the ***forced discrete Euler-Lagrange equations***:

$$D_1 L_d(q_{k+1}, q_{k+2}) + D_2 L_d(q_k, q_{k+1}) + F_d^-(q_{k+1}, q_{k+2}) + F_d^+(q_k, q_{k+1}) = 0. \tag{1.28}$$

The simplest example of discrete forces is to take

$$\begin{aligned} F_d^-(q_k, q_{k+1}) &= F(q_k) \\ F_d^+(q_k, q_{k+1}) &= 0, \end{aligned}$$

which, together with the discrete Lagrangian (1.3), gives the forced Euler-Lagrange equations

$$M\left(\frac{q_{k+1} - 2q_k + q_{k-1}}{h^2}\right) = -\nabla V(q_k) + F(q_k).$$

The position-momentum form of a variational integrator with forcing is useful for implementation purposes. This is given by

$$\begin{aligned} p_k &= -D_1 L_d(q_k, q_{k+1}) - F_d^-(q_k, q_{k+1}) \\ p_{k+1} &= D_2 L_d(q_k, q_{k+1}) + F_d^+(q_k, q_{k+1}). \end{aligned}$$

As an example of a variational integrator applied to a nonconservative system, in Figure 1.3 we plot the energy evolution of a Lagrangian system with dissipation added, which is simulated using a low-order variational integrator with forcing, as in (1.28), and a standard high-order Runge-Kutta method. Despite the disadvantage of being low-order, the variational method tracks the error decay more accurately as it does not artificially dissipate energy for stability purposes.

1.6 Conservation Properties of Variational Integrators

1.6.1 Noether's theorem and momentum conservation

One of the important features of variational systems is that symmetries of the system lead to momentum conservation laws of the Euler-Lagrange equations, a classical result known as Noether's theorem.

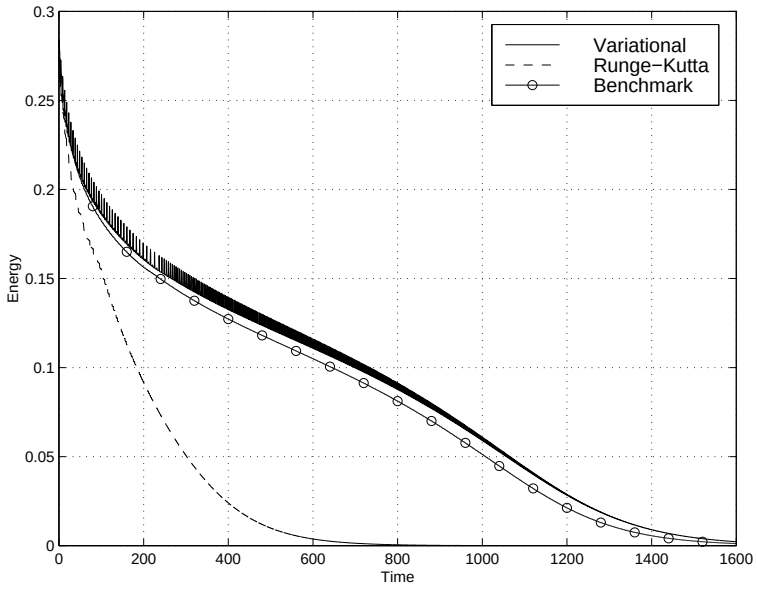


Figure 1.3: Energy evolution for a dissipative mechanical system, for a second-order variational integrator and a fourth-order Runge-Kutta method. The benchmark solution is a very expensive and accurate simulation. Observe that the variational method correctly captures the rate of decay of the energy, unlike the Runge-Kutta method.

Consider a one-parameter group of curves $q^\varepsilon(t)$, with $q^0(t) = q(t)$, which have the property that $L(q^\varepsilon(t), \dot{q}^\varepsilon(t)) = L(q(t), \dot{q}(t))$ for all ε. When the Lagrangian is invariant in this manner, then we have a symmetry of the system, and we write

$$\xi(t) = \left.\frac{\partial q^\varepsilon(t)}{\partial \varepsilon}\right|_{\varepsilon=0} \tag{1.29}$$

for the infinitesimal symmetry direction.

The fact that the Lagrangian is invariant means that the action integral is also invariant, so its derivative with respect to ε will be zero. If $q(t)$ is a solution trajectory, then we can set the Euler-Lagrange term in equation (1.1) to zero to obtain

$$0 = \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} \int_0^T L\big(q(t), \dot{q}(t)\big)\, dt = \frac{\partial L}{\partial \dot{q}}\big(q(T), \dot{q}(T)\big) \cdot \xi(T) - \frac{\partial L}{\partial \dot{q}}\big(q(0), \dot{q}(0)\big) \cdot \xi(0). \tag{1.30}$$

The terms on the right hand side above are the final and initial momentum in the direction ξ, which are thus equal. This is the statement of ***Noether's theorem***.

As examples, consider the one-parameter groups $q^\varepsilon(t) = q(t) + \varepsilon v$ and $q^\varepsilon(t) = \exp(\varepsilon\Omega)q(t)$ for any vector v and skew-symmetric matrix Ω. The transformations give translations and rotations, respectively, and evaluating (1.30) for these cases gives conservation of linear and angular momentum, assuming that the Lagrangian is indeed invariant under these transformations.

Geometric aside. More generally, we may consider an arbitrary Lie group G, with Lie algebra $\mathfrak{g}$, rather than the one-dimensional groups taken above. The analogue of $\xi(t)$ is then the infinitesimal generator $\xi_Q : Q \to TQ$, for any $\xi \in \mathfrak{g}$, corresponding to an action of G on Q whose lift to TQ leaves L invariant. Equation (1.30) then becomes $(\partial L/\partial \dot{q}) \cdot \xi_Q|_0^T = 0$, which means that the momentum map $J_L : TQ \to \mathfrak{g}^*$ is conserved, where $J_L(q, \dot{q}) \cdot \xi = (\partial L/\partial \dot{q}) \cdot \xi_Q(q)$. While we must generally take many one-parameter groups, such as translations by any vector v, to show that a quantity such as linear momentum is conserved, with this general framework we can take $\mathfrak{g}$ to be the space of all vs, and thus obtain conservation of linear momentum with only a single group, albeit multidimensional, as is done in §2.1.4.

1.6.2 Discrete time Noether's theorem and discrete momenta

A particularly nice feature of the variational derivation of momentum conservation is that we simutaneously derive both the expression for the conserved quantity and the theorem that it is conserved. By using the variational derivation in the discrete time case, we can thus obtain the definition of discrete time momenta, as well as a discrete time Noether's theorem implying that they are conserved.

Take a one-parameter group of discrete time curves $\{q_k^\varepsilon\}_{k=0}^N$, with $q_k^0 = q_k$, such that $L_d(q_k^\varepsilon, q_{k+1}^\varepsilon) = L_d(q_k, q_{k+1})$ for all ε and k. The infinitesimal symmetry for such an invariant discrete Lagrangian is written

$$\xi_k = \left.\frac{\partial q_k^\varepsilon}{\partial \varepsilon}\right|_{\varepsilon=0}. \tag{1.31}$$

Invariance of the discrete Lagrangian implies invariance of the action sum, and so its ε derivative will be zero. Assuming that $\{q_k\}$ is a solution trajectory, then (1.4) becomes

$$0 = \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} \sum_{k=0}^{N-1} L_d(q_k^\varepsilon, q_{k+1}^\varepsilon) = D_1 L_d(q_0, q_1) \cdot \delta\xi_0 + D_2 L_d(q_{N-1}, q_N) \cdot \delta\xi_N. \tag{1.32}$$

Observing that $0 = D_1 L_d(q_0, q_1) \cdot \xi_0 + D_2 L_d(q_0, q_1) \cdot \xi_1$ as L_d is invariant, we thus have the ***discrete Noether's theorem***

$$D_2 L_d(q_{N-1}, q_N) \cdot \xi = D_2 L_d(q_0, q_1) \cdot \xi, \tag{1.33}$$

where the discrete momentum in the direction ξ is given by $D_2 L_d(q_k, q_{k+1}) \cdot \xi$.

Consider the example discrete Lagrangian (1.13) with $\alpha = 0$, and assume that $q \in Q \equiv \mathbb{R}^3$ and that V is a function of the norm of q only. This is the case of a particle in a radial potential for example. Then the discrete Lagrangian is invariant under rotations $q_k^\varepsilon = \exp(\varepsilon\Omega) q_k$, for any

skew-symmetric matrix $\Omega \in \mathbb{R}^{3\times 3}$. Evaluating (1.33) in this case gives

$$q_N \times M\left(\frac{q_N - q_{N-1}}{t_N - t_{N-1}}\right) = q_1 \times M\left(\frac{q_1 - q_0}{t_1 - t_0}\right). \tag{1.34}$$

We have thus computed the correct expressions for the discrete angular momentum, and shown that it is conserved. Note that while this expression may seem obvious, in more complicated examples this will not be the case.

Geometric aside. As in the continuous case, we can extend the above derivation to multidimensional groups and define a full discrete momentum map $J_{L_d} : Q \times Q \to \mathfrak{g}^*$ by $J_L(q_0, q_1)\cdot\xi = D_2L_d(q_0, q_1)\cdot\xi_Q(q_1)$. In fact there are two discrete momentum maps, corresponding to D_1L_d and D_2L_d, but they are equal whenever L_d is invariant, as we shall see in §2.2.3.

1.6.3 Continuous time symplecticity

In addition to the conservation of energy and momenta, Lagrangian mechanical systems also conserve another quantity known as a symplectic bilinear form.

Consider a two-parameter set of initial conditions $(q_0^{\epsilon,\nu}, v_0^{\epsilon\nu})$ so that $(q^{\epsilon,\nu}(t), v^{\epsilon,\nu}(t))$ is the resulting trajectory of the system. The corresponding variations are denoted

$$\delta q_1^\epsilon(t) = \frac{\partial}{\partial\nu}q^{\epsilon,\nu}(t)\bigg|_{\nu=0} \qquad \delta q_2^\nu(t) = \frac{\partial}{\partial\epsilon}q^{\epsilon,\nu}(t)\bigg|_{\epsilon=0} \qquad \delta^2 q(t) = \frac{\partial}{\partial\epsilon}\frac{\partial}{\partial\nu}q^{\epsilon,\nu}(t)\bigg|_{\epsilon,\nu=0}$$

and we write $\delta q_1(t) = \delta q_1^0(t)$, $\delta q_2(t) = \delta q_2^0(t)$ and $q^\epsilon(t) = q^{\epsilon,0}(t)$. We now compute the second derivative of the action integral to be

$$\begin{aligned}
\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}\frac{\partial}{\partial\nu}\bigg|_{\nu=0} S(q^{\epsilon,\nu}) &= \frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}(DS(q^\epsilon)\cdot\delta q_1^\epsilon) \\
&= \frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}\left(\frac{\partial L}{\partial v^i}(\delta q_1^\varepsilon)^i(T) - \frac{\partial L}{\partial v^i}(\delta q_1^\varepsilon)^i(0)\right) \\
&= \frac{\partial^2 L}{\partial q^j\partial v^i}\delta q_1^i(T)\delta q_2^j(T) + \frac{\partial^2 L}{\partial v^j\partial v^i}\delta q_1^i(T)\delta\dot q_2^j(T) + \frac{\partial L}{\partial v^i}\delta^2 q^i(T) \\
&\quad - \frac{\partial^2 L}{\partial q^j\partial v^i}\delta q_1^i(0)\delta q_2^j(0) - \frac{\partial^2 L}{\partial v^j\partial v^i}\delta q_1^i(0)\delta\dot q_2^j(0) - \frac{\partial L}{\partial v^i}\delta^2 q^i(0).
\end{aligned}$$

Here and subsequently, repeated indices in a product indicate sum over the index range. If we reverse the order of differentiation with respect to ϵ and ν, then by symmetry of mixed partial derivatives

we will obtain an equivalent expression. Subtracting this from the above equation then gives

$$\frac{\partial^2 L}{\partial q^j \partial v^i}\left[\delta q_1^i(T)\delta q_2^j(T) - \delta q_2^i(T)\delta q_1^j(T)\right] + \frac{\partial^2 L}{\partial v^j \partial v^i}\left[\delta q_1^i(T)\delta \dot{q}_2^j(T) - \delta q_2^i(T)\delta \dot{q}_1^j(T)\right]$$
$$= \frac{\partial^2 L}{\partial q^j \partial v^i}\left[\delta q_1^i(0)\delta q_2^j(0) - \delta q_2^i(0)\delta q_1^j(0)\right] + \frac{\partial^2 L}{\partial v^j \partial v^i}\left[\delta q_1^i(0)\delta \dot{q}_2^j(0) - \delta q_2^i(0)\delta \dot{q}_1^j(0)\right].$$

Each side of this expression is an antisymmetric bilinear form evaluated on the variations δq_1 and δq_2. The fact that this evaluation gives the same result at $t = 0$ and at $t = T$ implies that the bilinear form itself is preserved by the Euler-Lagrange equations. This bilinear form is called the ***symplectic form*** of the system, and the fact that it is preserved is called symplecticity of the flow of the Euler-Lagrange equations.

This is a conservation property in the same way that momenta and energy are conservation properties of Lagrangian mechanical systems, and it has a number of important consequences. Examples of this include Liouville's theorem, which states that phase space volume is preserved by the time evolution of the system, and fourfold symmetry of the eigenvalues of linearizations of the system, so that if λ is an eigenvalue, so too are $-\lambda$, $\bar{\lambda}$ and $-\bar{\lambda}$. There are many other important examples, see Marsden and Ratiu [1999].

Geometric aside. The above derivation can be written using differential geometric notation as follows. The boundary terms in the action variation equation (1.1) are intrinsically given by $\Theta_L = (\mathbb{F}L)^*\Theta$, the pullback under the Legendre transform of the canonical one-form $\Theta = p_i dq^i$ on T^*Q. We thus have $dS = (F_L^t)^*\Theta_L - \Theta_L$ and so using $d^2 = 0$ (which is the intrinsic statement of symmetry of mixed partial derivatives) we obtain $0 = d^2S = (F_L^t)^*(d\Theta_L) - d\Theta_L$. The symplectic two-form above is thus $\Omega_L = -d\Theta_L$, and we recover the usual statement of symplecticity of the flow F_L^t for Lagrangian systems, as we shall show in greater detail in § 2.1.3.

1.6.4 Discrete time symplecticity

As we have seen above, symplecticity of continuous time Lagrangian systems is a consequence of the variational structure. There is thus an analogous property of discrete Lagrangian systems.

Consider a two-parameter set of initial conditions $(q_0^{\varepsilon,\nu}, q_1^{\varepsilon,\nu})$ and let $\{q_k^{\varepsilon,\nu}\}_{k=0}^N$ be the resulting discrete trajectory. We denote the corresponding variations by

$$\delta q_k^\varepsilon = \frac{\partial}{\partial \nu} q_k^{\varepsilon,\nu}\bigg|_{\nu=0} \qquad \delta \bar{q}_k^\nu = \frac{\partial}{\partial \varepsilon} q_k^{\varepsilon,\nu}\bigg|_{\varepsilon=0} \qquad \delta^2 q_k = \frac{\partial}{\partial \varepsilon}\frac{\partial}{\partial \nu} q_k^{\varepsilon,\nu}\bigg|_{\varepsilon,\nu=0}$$

and we write $\delta q_k = \delta q_k^0$, $\delta \bar{q}_k = \delta \bar{q}_k^0$ and $q_k^\varepsilon = q^{\varepsilon,0}$ for $k = 0, \ldots, N$. The second derivative of the

action sum is thus given by

$$\begin{aligned}\frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0}\frac{\partial}{\partial\nu}\bigg|_{\nu=0}S_d(\{q_k^{\varepsilon,\nu}\}) &= \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0}\left(DS_d(\{q_k^{\varepsilon}\})\cdot\delta q^{\varepsilon}\right)\\ &= \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0}\left(D_{1i}L_d(q_0^{\varepsilon},q_1^{\varepsilon})\,(\delta q_0^{\varepsilon})^i + D_{2i}L_d(q_{N-1}^{\varepsilon},q_N^{\varepsilon})\,(\delta q_N^{\varepsilon})^i\right)\\ &= D_{1j}D_{1i}L_d(q_0,q_1)\delta q_0^i\delta\bar{q}_0^j + D_{2j}D_{1i}L_d(q_0,q_1)\delta q_0^i\delta\bar{q}_1^j\\ &\quad + D_{1j}D_{2i}L_d(q_{N-1},q_N)\delta q_N^i\delta\bar{q}_{N-1}^j + D_{2j}D_{2i}L_d(q_{N-1},q_N)\delta q_N^i\delta\bar{q}_N^j\\ &\quad + D_{1i}L_d(q_0,q_1)\delta^2q_0^i + D_{2i}L_d(q_{N-1},q_N)\delta^2q_N^i. \qquad (1.35)\end{aligned}$$

By symmetry of mixed partial derivatives, reversing the order of differentiation above will give an equivalent expression. Subtracting one from the other will thus give zero, and rearranging the resulting equation we obtain

$$D_{1j}D_{2i}L_d(q_{N-1},q_N)\left[\delta q_N^i\delta\bar{q}_{N-1}^j - \delta\bar{q}_N^i\delta q_{N-1}^j\right] = D_{2j}D_{1i}L_d(q_0,q_1)\left[\delta\bar{q}_0^i\delta q_1^j - \delta q_0^i\delta\bar{q}_1^j\right]. \qquad (1.36)$$

We can also repeat the derivation in (1.35) and (1.36) for $L_d(q_0^{\varepsilon,\nu},q_1^{\varepsilon,\nu})$, rather than the entire action sum, to obtain

$$D_{2j}D_{1i}L_d(q_0,q_1)\left[\delta\bar{q}_0^i\delta q_1^j - \delta q_0^i\delta\bar{q}_1^j\right] = D_{1j}D_{2i}L_d(q_0,q_1)\left[\delta q_1^i\delta\bar{q}_0^j - \delta\bar{q}_1^i\delta q_0^j\right], \qquad (1.37)$$

which can also be directly seen from the symmetry of mixed partial derivatives. Substituting this into (1.36) now gives

$$D_{1j}D_{2i}L_d(q_{N-1},q_N)\left[\delta q_N^i\delta\bar{q}_{N-1}^j - \delta\bar{q}_N^i\delta q_{N-1}^j\right] \quad = \quad D_{1j}D_{2i}L_d(q_0,q_1)\left[\delta q_1^i\delta\bar{q}_0^j - \delta\bar{q}_1^i\delta q_0^j\right]. \qquad (1.38)$$

We can now see that each side of this equation is an antisymmetric bilinear form, which we call the ***discrete symplectic form***, evaluated on the variations δq_k and $\delta\bar{q}_k$. The two sides give this expression at the first time step and the final time step, so we have that the discrete symplectic form is preserved by the time evolution of the discrete system.

In the next section we will consider some numerical consequences of this property.

Geometric aside. Intrinsically we can identify two one-forms $\Theta_{L_d}^+ = D_2L_d dq_1$ and $\Theta_{L_d}^- = D_1L_d dq_0$, so that $dS_d = (F_{L_d}^N)^*\Theta_{L_d}^+ + \Theta_{L_d}^-$. Using $d^2 = 0$ (symmetry of mixed partial derivatives) gives $0 = d^2S_d = (F_{L_d}^N)^*(d\Theta_{L_d}^+) + d\Theta_{L_d}^-$ and so defining the discrete symplectic two-forms $\Omega_{L_d}^{\pm} = -d\Theta_{L_d}^{\pm}$ gives $(F_{L_d}^N)^*\Omega_{L_d}^+ = -\Omega_{L_d}^-$, which is the intrinsic form of (1.36). However, we observe that $0 = d^2L_d = d(\Theta_{L_d}^+ + \Theta_{L_d}^-) = -\Omega_{L_d}^+ - \Omega_{L_d}^-$ and hence $\Omega_{L_d}^+ = -\Omega_{L_d}^-$, which is (1.37). Combining this with our previous expression then gives $(F_{L_d}^N)^*\Omega_{L_d}^+ = \Omega_{L_d}^+$ as the intrinsic form of (1.38), discrete

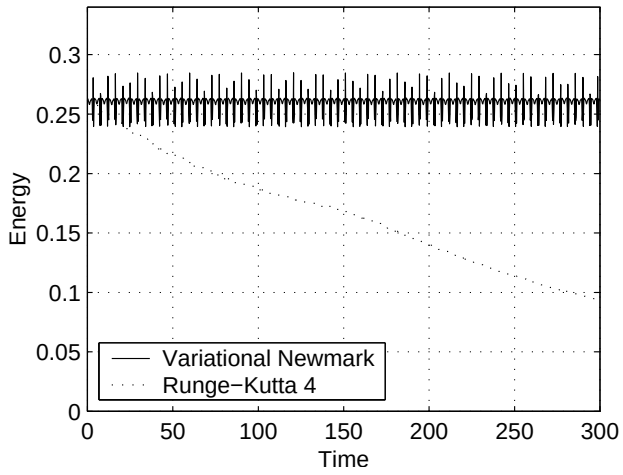


Figure 1.4: Energy computed with variational second-order Newmark and fourth-order Runge-Kutta. Note that the variational method does not artificially dissipate energy.

symplecticity of the evolution, as will be investigated in § 2.2.2.

Observe that using the discrete Legendre transforms we have $\Theta^{\pm}_{L_d} = (\mathbb{F}L_d^{\pm})^*\Theta$, where $\Theta = p_i dq^i$ is the canonical one-form on T^*Q. The expression (1.38) thus shows that the map $\mathbb{F}L_d^+ \circ F_{L_d} \circ (\mathbb{F}L_d^+)^{-1}$ preserves the canonical symplectic two-form Ω on T^*Q. Variational integrators are thus symplectic methods in the standard sense, a point which will be further discussed in §2.4.3 and §2.4.4.

1.6.5 Backward error analysis

We now briefly consider why preservation of a symplectic form may be advantageous numerically. We first consider a numerical example, and then the theory which explains it.

Approximate energy conservation. If we use a variational method to simulate a nonlinear model system and plot the energy versus time, then we obtain a graph like that in Figure 1.4. For comparison, this graph also shows the energy curve for a simulation with a standard method such as RK4 (the common fourth-order Runge-Kutta method).

The system being simulated here is purely conservative and so there should be no loss of energy over time. The striking aspect of this graph is that while the energy associated with a standard method decays due to numerical damping, for the Newmark method the energy error remains bounded. This may be understood by recognizing that the integrator is symplectic, that is, it preserves the same two-form on state space as the true system.

Backward error analysis. To understand the above numerics it is necessary to use the concept of backward error analysis, whereby we construct a ***modified system*** which is a model for the

numerical method. Here we only give a simple outline of the procedure; for details and proofs see Hairer et al. [2002] or Reich [1999a].

Consider an ODE $\dot{x} = f(x)$ and an integrator $x_{k+1} = F(x_k)$ which approximates the evolution of f. For a given initial condition x_0 let $x(t)$ be the true solution of $\dot{x} = f(x)$ and let $\{x_k\}_{k=0}^N$ be the discrete time approximation generated by F. Now let $\dot{\bar{x}} = \bar{f}(\bar{x})$ be a second ODE, called the ***modified system***,[5] for which the resulting trajectory $\bar{x}(t)$ exactly samples the points x_k, so that $\bar{x}(k\Delta t) = x_k$. We say that the ***backward error*** of F to f is the error $\|\bar{f} - f\|$, measured in an appropriate norm. This contrasts with the usual ***forward error*** of F to f given by $\|x_k - x(k\Delta t)\|$, which is also known as trajectory error.

We can now understand why variational integrators are different to standard methods. Namely, their modified systems are Lagrangian systems[6]. That is, for a given Lagrangian L, a variational integrator is exactly solving a system $\bar{L}$, which is close to L. This means that the discrete trajectory (which is exactly sampling the trajectory of $\bar{L}$) has all of the properties of a conservative mechanical system, such as energy conservation.

In the particular case of energy, we see in Figure 1.5a the phase space portrait of the exact solution of the nonlinear pendulum equations (a conservative Lagrangian system), together with the discrete time trajectories of a variational integrator and a standard Runge-Kutta method. Clearly the Runge-Kutta method is dissipative, and so its trajectory limits to the origin. In contrast, the variational integrator has an exactly conserved energy (the energy of the modified system $\bar{L}$) and so it remains on its level set for all time.

While it is not generally possible to calculate the modified system exactly, it is possible to find $\mathcal{O}(\Delta t)^r$ truncations of it. In Figure 1.5b the energy level set of the third-order truncation of the modified equation is plotted, and we see that indeed the variational integrator remains almost on it. This also explains why the energy plots for variational integrators contain a typical oscillation about the true energy. The modified energy level set will be close to the true energy level set everywhere, but it will typically be inside it at some locations and outside it at others. As the discrete trajectory evolves on the modified energy level set it thus has a non-modified energy which is sometimes greater than and sometimes less that the true value, but which never deviates very far. This results in the characteristic energy oscillation.

[5] In fact for most integrators F there will not be an exact modified system $\bar{f}$, but instead there will only be a system such that $\bar{x}(k\Delta t) = x_k + C_1 \exp(-C_2/(\Delta t))$, so the modified system is said to be ***exponentially close*** to the discrete time integrator. For Δt sufficiently small this is effectively as good as a true modified system. Furthermore, these estimates do not hold for all time, but rather for times on the order of $\exp(C_3/(\Delta t))$, which once again is essentially as good as forever for most computational purposes.

[6] This is typically done in the literature on the Hamiltonian side, by showing that symplectic integrators have Hamiltonian modified systems. As all variational integrators are symplectic and regular Hamiltonian systems are Lagrangian the result follows.

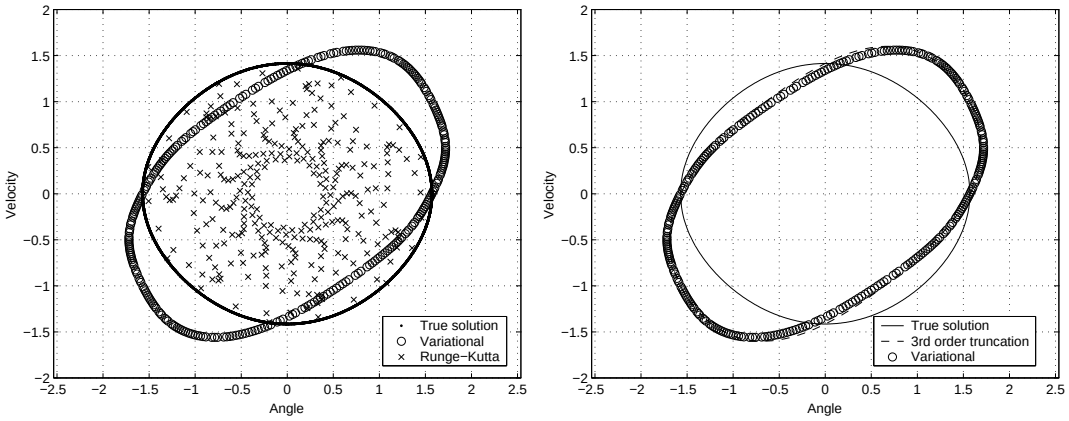


Figure 1.5: Phase space plots of: (a) symplectic method compared to non-symplectic Runge-Kutta method for the pendulum system, (b) third-order truncation of the modified energy.

1.7 Multisymplectic systems and variational integrators

There are two basic approaches with can be taken to extend the technique of variational integrators to PDEs. The first is to somehow discretize the spatial dimensions in such a way that the resulting semidiscrete problem is a Lagrangian ODE system to which variational integrators can be applied directly. The second is to use the variational technique to discretize both space and time simultaneously, which ensures that both the space–semidiscrete and time–semidiscrete systems will be Lagrangian, as well as opening the way to truly space-time discretizations such as that used for asynchronous time stepping in §6.

We focus here on the second approach, discretizing both space and time variationally, as it provides a more powerful framework for analyzing the properties of the resulting fully discrete system. The setting in which we formulate space-time variational systems is that of ***multisymplectic mechanics***, which we now briefly outline.

1.7.1 Variational multisymplectic mechanics

Geometry of multisymplectic systems. The ***base space*** $\mathcal{X}$ is the space of independent variables, typically space-time $(t, x^1, \ldots, x^n)$ with t being time and $x^1, \ldots, x^n$ being space dimensions. To simplify notation we will frequently treat time as the zeroth space coordinate, so that $x^0 = t$ and the independent variables are $(x^0, \ldots, x^n)$.

At each space-time point we have some number of dependent variables $q^1, \ldots, q^m$, which we regard as elements of a ***fiber*** over the space-time point. The set of all space-time points together with the independent fields at each point is thus a fiber bundle Y over $\mathcal{X}$, known as the ***configuration bundle***. A ***section*** of this bundle is a map $q : \mathcal{X} \to Y$ which leaves the base space point fixed, and thus gives independent variable values for any choice of dependent variables. We will typically write

a section as $q^i = q^i(t, x^1, \ldots, ^n)$ for $i = 1, \ldots, m$, and we picture it as a hypersurface in Y, as shown in Figure 1.6.

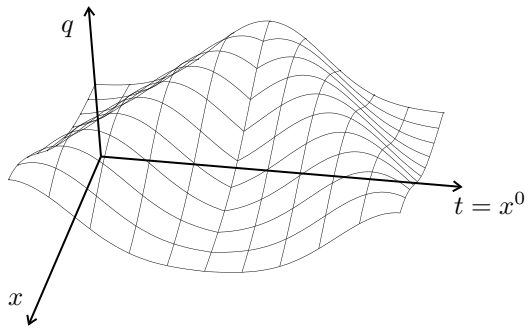


Figure 1.6: A graphical representation of a section of a fiber bundle, showing how a section gives the independent variables q at each space-time point (t, x).

Just as the configuration bundle Y corresponds to the configuration manifold Q of traditional mechanics (together with time), we require the space-time analog of the tangent bundle TQ. This is the ***first jet bundle*** J^1Y, which is the space Y together with the space of derivatives of the independent variables q. We can thus write points in J^1Y as $(x^0, \ldots, x^n, y^1, \ldots, y^m, v^1{}_0, \ldots, v^m{}_n)$, where $v^i{}_j$ corresponds to the derivative of q^i with respect to x^j, for $i = 1, \ldots, m$ and $j = 0, \ldots, n$. If we regard J^1Y as a fiber bundle over $\mathcal{X}$, then a section $q(x^0, \ldots, x^n)$ of Y given a section $j^1q(x^0, \ldots, x^n)$ of J^1Y by

$$j^1q(x^0, \ldots, x^n) = \left(x^0, \ldots, x^n, q^1, \ldots, q^m, q^1{}_{,1} = \frac{\partial q^1}{\partial x^1}, \ldots, q^m{}_n \frac{\partial q^m}{\partial x^n}\right). \tag{1.39}$$

Dynamics of multisymplectic systems. To describe the dynamics of a multisymplectic system, we proceed in a similar fashion to standard mechanics and define a ***Lagrangian*** $L : J^1Y \to \mathbb{R}$, which is now a function of a space-time point, the independent variables, and their derivatives. We then integrate the Lagrangian over space-time to find the ***action***

$$S(q) = \int_{\mathcal{X}} L\big(j^1q(x^0, \ldots, x^n)\big)\, dx^0 \ldots dx^n. \tag{1.40}$$

We then look for configurations q which are stationary points of the action, holding q fixed on the boundary of $\mathcal{X}$. The action variation is then

$$\begin{aligned} \delta S(q) &= \int_{\mathcal{X}} \left[\frac{\partial L}{\partial q^i}\delta q^i + \frac{\partial L}{\partial v^i{}_j}\delta q^i{}_{,j}\right] dx^0 \ldots dx^n \\ &= \int_{\mathcal{X}} \left[\frac{\partial L}{\partial q^i} - \frac{d}{dx^j}\frac{\partial L}{\partial v^i{}_j}\right] \delta q^i\, dx^0 \ldots dx^n + \int_{\partial\mathcal{X}} \frac{\partial L}{\partial v^i{}_j}\delta q^i\, dx^0 \ldots dx^n, \end{aligned} \tag{1.41}$$

where $i = 0, \ldots, m$, $j = 1, \ldots, n$ and we have used integration by parts. Taking the variations δq^i to be zero on the boundary $\partial\mathcal{X}$ of space-time eliminates the second term in (1.41) and requiring that the remainder be zero for all such δq^i gives the ***Euler-Lagrange*** equations

$$\frac{\partial L}{\partial q^i}(x^0, \ldots, x^n) - \frac{d}{dx^j}\left(\frac{\partial L}{\partial v^i{}_j}(x^0, \ldots, x^n)\right) = 0. \tag{1.42}$$

We see here that the action variation (1.41) gives the classical *weak form* of the PDE describing the system, while the Euler-Lagrange equations give the corresponding *strong form*.

Multisymplectic systems have a great deal of interesting geometry associated with them, as discussed in Gotay, Isenberg, and Marsden [1997], Marsden et al. [1998], and Marsden and Shkoller [1999], for example. In particular, the variational structure ensures certain conservation properties for multisymplectic systems, so that they have a Noether's theorem describing the conserved momenta arising from symmetries, and they have an extension of the symplectic structure of Lagrangian mechanics (hence the name "multisymplectic"). These points are discussed in detail in §5.

Nonlinear wave equation example. Consider a system with two independent variables (t, x) and one dependent variable q. The base space is thus $\mathcal{X} = \mathbb{R}^2$, the configuration bundle is $Y = \mathbb{R}^2 \times \mathbb{R}$, and so the first jet bundle is $J^1Y = \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}^2$, with coordinates (t, x, q, v_t, v_x). Taking the Lagrangian to be

$$L(t, x, q, v_t, v_x) = \frac{1}{2}v_t^2 - \frac{c^2}{2}v_x^2 + F(q), \tag{1.43}$$

for some potential $F(q)$, we can readily calculate that the Euler-Lagrange equations (1.42) are the regular nonlinear wave equation

$$q_{,tt} = c^2 q_{,xx} + F'(q) \tag{1.44}$$

with wavespeed c, where F' is the derivative of F.

1.7.2 Multisymplectic discretizations

We now consider variational discretizations of multisymplectic systems, as introduced by Marsden et al. [1998]. For simplicity we restrict outselves here to the case of one space dimension and one time dimension.

Just as with variational integrators for ODEs we discretize space and time, for example by taking the rectangular grid in Figure 1.7 with step sizes Δt and Δx in time and space. The space-time nodes are thus (t_i, x_j), at the i-th time step and j-th spatial step, and we write $q_{i,j}$ for the value of the field variable at the node.

To approximate the dynamics of a continuous Lagrangian multisymplectic system, we need to

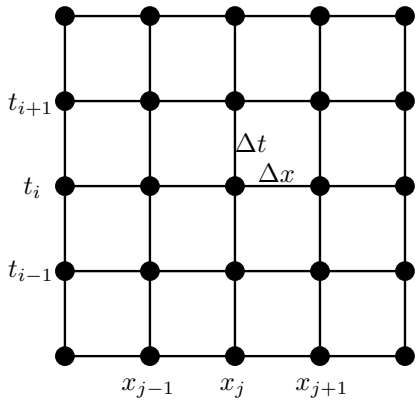


Figure 1.7: An example mesh for a multisymplectic space-time discretization. Here Δt and Δx are the time and space step sizes, respectively, space is horizontal and time is vertical.

approximate the action integral over a single space-time element with a ***discrete Lagrangian***

$$L_d(q_{i,j}, q_{i+1,j}, q_{i,j+1}, q_{i+1,j+1}) \approx \int_{[t_i,t_{i+1}]\times[x_j,x_{j+1}]} L(j^1 q(t,x))\, dt\, dx, \tag{1.45}$$

where $q(t,x)$ is a continuous function which approximates $q_{i,j}$, with $q(t_i, x_j) \approx q_{i,j}$.

To approximate the complete action integral (1.40) we now sum over all elements to obtain the discrete action

$$S_d(q) = \sum_{i=0}^{k-1} \sum_{j=0}^{\ell-1} L_d(q_{i,j}, q_{i+1,j}, q_{i,j+1}, q_{i+1,j+1}), \tag{1.46}$$

and we require that this be stationary with respect to variations in $q_{i,j}$ for each node (t_i, x_j) which is in the interior of the space-time region. Taking such a representative (t_i, x_j), we see that there are four terms in (1.46) which contain $q_{i,j}$, and so stationarity gives the ***discrete Euler-Lagrange equations***

$$\begin{aligned} &D_1 L_d(q_{i,j}, q_{i+1,j}, q_{i,j+1}, q_{i+1,j+1}) + D_2 L_d(q_{i-1,j}, q_{i,j}, q_{i-1,j+1}, q_{i,j+1}) \\ &\qquad + D_3 L_d(q_{i,j-1}, q_{i+1,j-1}, q_{i,j}, q_{i+1,j}) + D_4 L_d(q_{i-1,j-1}, q_{i,j-1}, q_{i-1,j}, q_{i,j}) = 0. \end{aligned} \tag{1.47}$$

Given initial data, such as $q_{i,j}$ for $i = 0$ and $i = 1$ and all j, these equations then define an integrator for calculating the values at time step $i = 2$, and hence advancing forward in time.

Just as with variational integrators for ODEs, the variational method of derivation for multi-symplectic discretizations means that the resulting numerical methods preserve (multi-)symplectic structures, have conserved momenta, and have excellent energy behavior. We discuss this in more detail in §6.1.4 and §6.3.

Wave equation example. For the wave equation Lagrangian (1.43), a simple approximation is given by

$$L_d(q_{i,j}, q_{i+1,j}, q_{i,j+1}, q_{i+1,j+1}) = (\Delta t)(\Delta x)\left[\frac{1}{2}\left(\frac{q_{i+1,j} - q_{i,j}}{\Delta t}\right)^2 - \frac{c^2}{2}\left(\frac{q_{i,j+1} - q_{i,j}}{\Delta t}\right)^2 + F(q_{i,j})\right]. \tag{1.48}$$

Computing the discrete Euler-Lagrange equations (1.47) for this Lagrangian gives

$$\frac{q_{i+1,j} - 2q_{i,j} + q_{i-1,j}}{(\Delta t)^2} - c^2\frac{q_{i,j+1} - 2q_{i,j} + q_{i,j-1}}{(\Delta x)^2} - F'(q_{i,j}) = 0, \tag{1.49}$$

which is clearly a discretization of (1.44).

Chapter 2

Discrete variational mechanics

2.1 Background: Lagrangian mechanics

2.1.1 Basic definitions

Consider a ***configuration manifold*** Q, with associated ***state space*** given by the tangent bundle TQ, and a ***Lagrangian*** $L : TQ \to \mathbb{R}$.

Given an interval $[0, T]$, define the ***path space*** to be

$$\mathcal{C}(Q) = \mathcal{C}([0, T], Q) = \{q \colon [0, T] \to Q \mid q \text{ is a } C^2 \text{ curve}\}$$

and the ***action map*** $\mathfrak{G} : \mathcal{C}(Q) \to \mathbb{R}$ to be

$$\mathfrak{G}(q) \equiv \int_0^T L(q(t), \dot{q}(t))dt. \tag{2.1}$$

It can be proved that $\mathcal{C}(Q)$ is a smooth manifold [Abraham, Marsden, and Ratiu, 1988], and $\mathfrak{G}$ is as smooth as L.

The tangent space $T_q\mathcal{C}(Q)$ to $\mathcal{C}(Q)$ at the point q is the set of C^2 maps $v_q : [0, T] \to TQ$ such that $\pi_Q \circ v_q = q$, where $\pi_Q : TQ \to Q$ is the canonical projection.

Define the ***second-order submanifold*** of $T(TQ)$ to be

$$\ddot{Q} \equiv \{w \in T(TQ) \mid T\pi_Q(w) = \pi_{TQ}(w)\} \subset T(TQ)$$

where $\pi_{TQ} : T(TQ) \to TQ$ and $\pi_Q : TQ \to Q$ are the canonical projections. $\ddot{Q}$ is simply the set of second derivatives $d^2q/dt^2(0)$ of curves $q : \mathbb{R} \to Q$, which are elements of the form $((q, \dot{q}), (\dot{q}, \ddot{q})) \in T(TQ)$.

Theorem 2.1. *Given a C^k Lagrangian L, $k \geq 2$, there exists a unique C^{k-2} mapping $D_{\mathrm{EL}}L : \ddot{Q} \to T^*Q$ and a unique C^{k-1} one-form Θ_L on TQ, such that, for all variations $\delta q \in T_q\mathcal{C}(Q)$ of $q(t)$, we*

have

$$\mathbf{d}\mathfrak{G}(q)\cdot\delta q=\int_0^T D_{\mathrm{EL}}L(\ddot{q})\cdot\delta q dt+\Theta_L(\dot{q})\cdot\hat{\delta q}\Big|_0^T, \tag{2.2}$$

where

$$\hat{\delta q}(t)=\left(\left(q(t),\frac{\partial q}{\partial t}(t)\right),\left(\delta q(t),\frac{\partial \delta q}{\partial t}(t)\right)\right).$$

The mapping $D_{\mathrm{EL}}L$ is called the **Euler-Lagrange map** *and has the coordinate expression*

$$\left(D_{\mathrm{EL}}L\right)_i=\frac{\partial L}{\partial q^i}-\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}.$$

The one-form Θ_L is called the **Lagrangian one-form** *and in coordinates is given by*

$$\Theta_L=\frac{\partial L}{\partial \dot{q}^i}\mathbf{d}q^i. \tag{2.3}$$

Proof. Computing the variation of the action map gives

$$\begin{aligned}\mathbf{d}\mathfrak{G}(q)\cdot\delta q&=\int_0^T\left[\frac{\partial L}{\partial q^i}\delta q^i+\frac{\partial L}{\partial \dot{q}^i}\frac{d}{dt}\delta q^i\right]dt\\&=\int_0^T\left[\frac{\partial L}{\partial q^i}-\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right]\cdot\delta q^i dt+\left[\frac{\partial L}{\partial \dot{q}^i}\delta q^i\right]_0^T\end{aligned}$$

using integration by parts, and the terms of this expression can be identified as the Euler-Lagrange map and the Lagrangian one-form. □

2.1.2 Lagrangian vector fields and flows

The ***Lagrangian vector field*** $X_L:TQ\to T(TQ)$ is a second-order vector field on TQ satisfying

$$D_{\mathrm{EL}}L\circ X_L=0 \tag{2.4}$$

and the ***Lagrangian flow*** $F_L:TQ\times\mathbb{R}\to TQ$ is the flow of X_L (we shall ignore issues related to global versus local flows, which are easily dealt with by restricting the domains of the flows). We shall write $F_L^t:TQ\to TQ$ for the map F_L at the frozen time t.

For an arbitrary Lagrangian, equation (2.4) may not uniquely define the vector field X_L and hence the flow map F_L may not exist. For now we will assume that L is such that these objects exist and are unique, and in §2.3.3 we will see under what conditions this is true.

A curve $q\in\mathcal{C}(Q)$ is said to be a ***solution of the Euler-Lagrange equations*** if the first term on the right-hand side of (2.2) vanishes for all variations $\delta q\in T_q\mathcal{C}(Q)$. This is equivalent to $(q,\dot{q})$

being an integral curve of X_L, and means that q must satisfy the ***Euler-Lagrange equations***

$$\frac{\partial L}{\partial q^i}(q,\dot{q}) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}(q,\dot{q})\right) = 0 \tag{2.5}$$

for all $t \in (0,T)$.

2.1.3 Lagrangian flows are symplectic

Define the ***solution space*** $\mathcal{C}_L(Q) \subset \mathcal{C}(Q)$ to be the set of solutions of the Euler-Lagrange equations. As an element $q \in \mathcal{C}_L(Q)$ is an integral curve of X_L, it is uniquely determined by the initial condition $(q(0),\dot{q}(0)) \in TQ$ and we can thus identify $\mathcal{C}_L(Q)$ with the space of initial conditions TQ.

Defining the ***restricted action map*** $\hat{\mathfrak{G}} : TQ \to \mathbb{R}$ to be

$$\hat{\mathfrak{G}}(v_q) = \mathfrak{G}(q), \quad q \in \mathcal{C}_L(Q) \text{ and } (q(0),\dot{q}(0)) = v_q,$$

we see that (2.2) reduces to

$$\begin{aligned}\mathbf{d}\hat{\mathfrak{G}}(v_q)\cdot w_{v_q} &= \Theta_L(\dot{q}(T))((F_L^T)_*(w_{v_q})) - \Theta_L(v_q)(w_{v_q}) \\ &= ((F_L^T)^*(\Theta_L))(v_q)(w_{v_q}) - \Theta_L(v_q)(w_{v_q})\end{aligned} \tag{2.6}$$

for all $w_{v_q} \in T_{v_q}(TQ)$. Taking a further derivative of this expression, and using the fact that $\mathbf{d}^2\hat{\mathfrak{G}} = 0$, we obtain

$$(F_L^T)^*(\Omega_L) = \Omega_L,$$

where $\Omega_L = \mathbf{d}\Theta_L$ is the ***Lagrangian symplectic form***, given in coordinates by

$$\Omega_L(q,\dot{q}) = \frac{\partial^2 L}{\partial q^i \partial \dot{q}^j}\mathbf{d}q^i \wedge \mathbf{d}q^j + \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}\mathbf{d}\dot{q}^i \wedge \mathbf{d}q^j.$$

2.1.4 Lagrangian flows preserve momentum maps

Suppose that a Lie group G, with Lie algebra $\mathfrak{g}$, acts on Q by the (left or right) action $\Phi : G \times Q \to Q$. Consider the tangent lift of this action to $\Phi^{TQ} : G \times TQ \to TQ$ given by $\Phi_g^{TQ}(v_q) = T(\Phi_g)\cdot v_q$, which is

$$\Phi^{TQ}\big(g,(q,\dot{q})\big) = \left(\Phi^i(g,q), \frac{\partial \Phi^i}{\partial q^j}(g,q)\,\dot{q}^j\right).$$

For $\xi \in \mathfrak{g}$ define the ***infinitesimal generators*** $\xi_Q : Q \to TQ$ and $\xi_{TQ} : TQ \to T(TQ)$ by

$$\xi_Q(q) = \frac{d}{dg}\Big(\Phi_g(q)\Big) \cdot \xi,$$
$$\xi_{TQ}(v_q) = \frac{d}{dg}\Big(\Phi_g^{TQ}(v_q)\Big) \cdot \xi.$$

In coordinates these are given by

$$\xi_Q(q) = \left(q^i, \frac{\partial \Phi^i}{\partial g^m}(e,q)\xi^m\right),$$
$$\xi_{TQ}(q,\dot{q}) = \left(q^i, \dot{q}^i, \frac{\partial \Phi^i}{\partial g^m}(e,q)\xi^m, \frac{\partial^2 \Phi^i}{\partial g^m \partial q^j}(e,q)\dot{q}^j\xi^m\right).$$

We now define the ***Lagrangian momentum map*** $J_L : TQ \to \mathfrak{g}^*$ to be

$$J_L(v_q) \cdot \xi = \Theta_L \cdot \xi_{TQ}(v_q). \tag{2.7}$$

It can be checked that an equivalent expression for J_L is

$$J_L(v_q) \cdot \xi = \left\langle \frac{\partial L}{\partial \dot{q}}, \xi_Q(q) \right\rangle,$$

where $\partial L/\partial \dot{q}$ represents the Legendre transformation, discussed shortly. This equation is convenient for computing momentum maps in examples: see Marsden and Ratiu [1999].

The traditional linear and angular momenta are momentum maps, with the linear momentum $J_L : T\mathbb{R}^n \to \mathbb{R}^n$ arising from the additive action of $\mathbb{R}^n$ on itself, and the angular momentum $J_L : T\mathbb{R}^n \to \mathfrak{so}(n)^*$ coming from the action of $SO(n)$ on $\mathbb{R}^n$.

An important property of momentum maps is ***equivariance***, which is the condition that the following diagram commutes.

$$\begin{array}{ccc} TQ & \xrightarrow{J_L} & \mathfrak{g}^* \\ {\scriptstyle \Phi_g^{TQ}}\downarrow & & \downarrow{\scriptstyle \mathrm{Ad}^*_{g^{-1}}} \\ TQ & \xrightarrow[J_L]{} & \mathfrak{g}^* \end{array} \tag{2.8}$$

In general, Lagrangian momentum maps are not equivariant, but we give here a simple sufficient condition for this property to be satisfied. Recall that a map $f : TQ \to TQ$ is said to be ***symplectic*** if $f^*\Omega_L = \Omega_L$. If, furthermore, f is such that $f^*\Theta_L = \Theta_L$, then f is said to be a ***special symplectic map***. Clearly a special symplectic map is also symplectic, but the converse does not hold.

Theorem 2.2. *Consider a Lagrangian system $L : TQ \to \mathbb{R}$ with a left action $\Phi : G \times Q \to Q$. If the lifted action $\Phi^{TQ} : G \times TQ \to TQ$ acts by special canonical transformations, then the Lagrangian momentum map $J_L : TQ \to \mathfrak{g}^*$ is equivariant.*

Proof. Observing that $(\Phi_g^{TQ})^{-1} = \Phi_{g^{-1}}^{TQ}$, we see that equivariance is equivalent to

$$J_L(v_q) \cdot \xi = J_L \circ T\Phi_{g^{-1}}(v_q) \cdot \mathrm{Ad}_{g^{-1}} \xi.$$

We now compute the right-hand side of this expression to give

$$\begin{aligned} J_L \circ \Phi_{g^{-1}}^{TQ}(v_q) \cdot \mathrm{Ad}_{g^{-1}} \xi &= \big\langle \Theta_L\big(\Phi_{g^{-1}}^{TQ}(v_q)\big), (\mathrm{Ad}_{g^{-1}} \xi)_{TQ}\big(\Phi_{g^{-1}}^{TQ}(v_q)\big)\big\rangle \\ &= \big\langle \Theta_L\big(\Phi_{g^{-1}}^{TQ}(v_q)\big), T(\Phi_{g^{-1}}^{TQ}) \cdot \xi_{TQ}(v_q)\big\rangle \\ &= \big\langle \big((\Phi_{g^{-1}}^{TQ})^* \Theta_L\big)(v_q), \xi_{TQ}(v_q)\big\rangle \\ &= \big\langle \Theta_L(v_q), \xi_{TQ}(v_q)\big\rangle, \end{aligned}$$

which is just $J_L(v_q) \cdot \xi$, as desired. Here we used the identity $(\mathrm{Ad}_g \xi)_M = \Phi_{g^{-1}}^* \xi_M$ [Marsden and Ratiu, 1999] to pass from the first to the second line. □

A Lagrangian $L : TQ \to \mathbb{R}$ is said to be ***invariant*** under the lift of the action $\Phi : G \times Q \to Q$ if $L \circ \Phi_g^{TQ} = L$ for all $g \in G$, and in this case the group action is said to be a ***symmetry*** of the Lagrangian. Differentiating this expression implies that the Lagrangian is ***infinitesimally invariant***, which is the statement $\mathbf{d}L \cdot \xi_{TQ} = 0$ for all $\xi \in \mathfrak{g}$.

Observe that if L is invariant, then this implies that Φ^{TQ} acts by special symplectic transformations, and so the Lagrangian momentum map is equivariant. To see this, we write $L \circ \Phi_g^{TQ} = L$ in coordinates to obtain $L(\Phi_g(q), \partial_q \Phi_g(q) \cdot \dot{q}) = L(q, \dot{q})$, and now differentiating this with respect to $\dot{q}$ in the direction δq gives

$$\frac{\partial L}{\partial \dot{q}}\big(\Phi_g(q), \partial_q \Phi_g(q) \cdot \dot{q}\big) \cdot \partial_q \Phi_g(q) \cdot \delta q = \frac{\partial L}{\partial \dot{q}}(q, \dot{q}) \cdot \delta q.$$

But the left- and right-hand sides are simply $(\Phi_g^{TQ})^* \Theta_L$ and Θ_L, respectively, evaluated on $((q, \dot{q}), (\delta q, \delta \dot{q}))$, and thus we have $(\Phi_g^{TQ})^* \Theta_L = \Theta_L$.

We will now show that, when the group action is a symmetry of the Lagrangian, then the momentum maps are preserved by the Lagrangian flow. This result was originally due to Noether [1918], using a technique similar to the one given below.

Theorem 2.3 (Noether's theorem). *Consider a Lagrangian system $L : TQ \to \mathbb{R}$ which is invariant under the lift of the (left or right) action $\Phi : G \times Q \to Q$. Then the corresponding Lagrangian momentum map $J_L : TQ \to \mathfrak{g}^*$ is a conserved quantity of the flow, so that $J_L \circ F_L^t = J_L$ for all times t.*

Proof. The action of G on Q induces an action of G on the space of paths $\mathcal{C}(Q)$ by pointwise action, so that $\Phi_g : \mathcal{C}(Q) \to \mathcal{C}(Q)$ is given by $\Phi_g(q)(t) = \Phi_g(q(t))$. As the action is just the integral of the

Lagrangian, invariance of L implies invariance of $\mathfrak{G}$ and the differential of this gives

$$\mathbf{d}\mathfrak{G}(q) \cdot \xi_{\mathcal{C}(Q)}(q) = \int_0^T \mathbf{d}L \cdot \xi_{TQ} dt = 0.$$

Invariance of $\mathfrak{G}$ also implies that Φ_g maps solution curves to solution curves and thus $\xi_{\mathcal{C}(Q)}(q) \in T_q\mathcal{C}_L$, which is the corresponding infinitesimal version. We can thus restrict $\mathbf{d}\mathfrak{G} \cdot \xi_{\mathcal{C}(Q)}$ to the space of solutions $\mathcal{C}_L$ to obtain

$$0 = \hat{\mathfrak{G}}(v_q) \cdot \xi_{TQ}(v_q) = \Theta_L(\dot{q}(T)) \cdot \xi_{TQ}(\dot{q}(T)) - \Theta_L(v_q) \cdot \xi_{TQ}(v_q).$$

Substituting in the definition of the Lagrangian momentum map J_L, however, shows that this is just $0 = J_L(F_L^T(v_q)) \cdot \xi - J_L(v_q) \cdot \xi$, which gives the desired result. □

We have thus seen that conservation of momentum maps is a direct consequence of the invariance of the variational principle under a symmetry action. The fact that the symmetry maps solution curves to solution curves will extend directly to discrete mechanics.

In fact, only infinitesimal invariance is needed for the momentum map to be conserved by the Lagrangian flow, as a careful reading of the above proof will show. This is because it is only necessary that the Lagrangian be invariant in a neighbourhood of a given trajectory, and so the global statement of invariance is stronger than necessary.

2.2 Discrete variational mechanics: Lagrangian viewpoint

Take again a configuration manifold Q, but now define the ***discrete state space*** to be $Q \times Q$. This contains the same amount of information as (is locally isomorphic to) TQ. A ***discrete Lagrangian*** is a function $L_d : Q \times Q \to \mathbb{R}$.

To relate discrete and continuous mechanics it is necessary to introduce a ***time step*** $h \in \mathbb{R}$, and to take L_d to depend on this time step. For the moment, we will take $L_d : Q \times Q \times \mathbb{R} \to \mathbb{R}$, and will neglect the h dependence except where it is important. We shall come back to this point later when we discuss the context of time-dependent mechanics and adaptive algorithms. However, the idea behind this was explained in the introduction.

Construct the increasing sequence of times $\{t_k = kh \mid k = 0, \ldots, N\} \subset \mathbb{R}$ from the time step h, and define the ***discrete path space*** to be

$$\mathcal{C}_d(Q) = \mathcal{C}_d(\{t_k\}_{k=0}^N, Q) = \{q_d \colon \{t_k\}_{k=0}^N \to Q\}.$$

We will identify a discrete trajectory $q_d \in \mathcal{C}_d(Q)$ with its image $q_d = \{q_k\}_{k=0}^N$, where $q_k = q_d(t_k)$.

The ***discrete action map*** $\mathfrak{G}_d : \mathcal{C}_d(Q) \to \mathbb{R}$ is defined by

$$\mathfrak{G}_d(q_d) = \sum_{k=0}^{N-1} L_d(q_k, q_{k+1}).$$

As the discrete path space $\mathcal{C}_d$ is isomorphic to $Q \times \cdots \times Q$ ($N+1$ copies), it can be given a smooth product manifold structure. The discrete action $\mathfrak{G}_d$ clearly inherits the smoothness of the discrete Lagrangian L_d.

The tangent space $T_{q_d}\mathcal{C}_d(Q)$ to $\mathcal{C}_d(Q)$ at q_d is the set of maps $v_{q_d} : \{t_k\}_{k=0}^N \to TQ$ such that $\pi_Q \circ v_{q_d} = q_d$, which we will denote by $v_{q_d} = \{(q_k, v_k)\}_{k=0}^N$.

The discrete object corresponding to $T(TQ)$ is the set $(Q \times Q) \times (Q \times Q)$. We define the ***projection operator*** π and the ***translation operator*** σ to be

$$\pi : ((q_0, q_1), (q_0', q_1')) \mapsto (q_0, q_1),$$
$$\sigma : ((q_0, q_1), (q_0', q_1')) \mapsto (q_0', q_1').$$

The ***discrete second-order submanifold*** of $(Q \times Q) \times (Q \times Q)$ is defined to be

$$\ddot{Q}_d \equiv \{w_d \in (Q \times Q) \times (Q \times Q) \mid \pi_1 \circ \sigma(w_d) = \pi_2 \circ \pi(w_d)\},$$

which has the same information content as (is locally isomorphic to) $\ddot{Q}$. Concretely, the discrete second-order submanifold is the set of pairs of the form $((q_0, q_1), (q_1, q_2))$.

Theorem 2.4. *Given a C^k discrete Lagrangian L_d, $k \geq 1$, there exists a unique C^{k-1} mapping $D_{\mathrm{DEL}}L_d : \ddot{Q}_d \to T^*Q$ and unique C^{k-1} one-forms $\Theta^+_{L_d}$ and $\Theta^-_{L_d}$ on $Q \times Q$, such that, for all variations $\delta q_d \in T_{q_d}\mathcal{C}(Q)$ of q_d, we have*

$$\mathbf{d}\mathfrak{G}_d(q_d) \cdot \delta q_d = \sum_{k=1}^{N-1} D_{\mathrm{DEL}}L_d((q_{k-1}, q_k), (q_k, q_{k+1})) \cdot \delta q_k + \Theta^+_{L_d}(q_{N-1}, q_N) \cdot (\delta q_{N-1}, \delta q_N) - \Theta^-_{L_d}(q_0, q_1) \cdot (\delta q_0, \delta q_1). \quad (2.9)$$

The mapping $D_{\mathrm{DEL}}L_d$ is called the **discrete Euler-Lagrange map** *and has coordinate expression*

$$D_{\mathrm{DEL}}L_d((q_{k-1}, q_k), (q_k, q_{k+1})) = D_2L_d(q_{k-1}, q_k) + D_1L_d(q_k, q_{k+1}).$$

The one-forms $\Theta^+_{L_d}$ and $\Theta^-_{L_d}$ are called the **discrete Lagrangian one-forms** *and in coordinates*

are

$$\Theta^+_{L_d}(q_0,q_1) = D_2L_d(q_0,q_1)\mathbf{d}q_1 = \frac{\partial L_d}{\partial q_1^i}dq_1^i, \tag{2.10a}$$

$$\Theta^-_{L_d}(q_0,q_1) = -D_1L_d(q_0,q_1)\mathbf{d}q_0 = -\frac{\partial L_d}{\partial q_0^i}dq_0^i. \tag{2.10b}$$

Proof. Computing the derivative of the discrete action map gives

$$\begin{aligned}\mathbf{d}\mathfrak{G}_d(q_d)\cdot\delta q_d &= \sum_{k=0}^{N-1}\left[D_1L_d(q_k,q_{k+1})\cdot\delta q_k + D_2L_d(q_k,q_{k+1})\cdot\delta q_{k+1}\right]\\ &= \sum_{k=1}^{N-1}\left[D_1L_d(q_k,q_{k+1}) + D_2L_d(q_{k-1},q_k)\right]\cdot\delta q_k\\ &\quad + D_1L_d(q_0,q_1)\cdot\delta q_0 + D_2L_d(q_{N-1},q_N)\cdot\delta q_N\end{aligned}$$

using a discrete integration by parts (rearrangement of the summation). Identifying the terms with the discrete Euler-Lagrange map and the discrete Lagrangian one-forms now gives the desired result. □

Unlike the continuous case, in the discrete case there are two one-forms that arise from the boundary terms. Observe, however, that $\mathbf{d}L_d = \Theta^+_{L_d} - \Theta^-_{L_d}$ and so using $\mathbf{d}^2 = 0$ shows that

$$\mathbf{d}\Theta^+_{L_d} = \mathbf{d}\Theta^-_{L_d}.$$

This will be reflected below in the fact that there is only a single discrete two-form, which is the same as the continuous situation and is important for symplecticity.

2.2.1 Discrete Lagrangian evolution operator and mappings

A ***discrete evolution operator*** X plays the same role as a continuous vector field, and is defined to be any map $X : Q\times Q \to (Q\times Q)\times(Q\times Q)$ satisfying $\pi\circ X = id$. The discrete object corresponding to the flow is the ***discrete map*** $F : Q\times Q \to Q\times Q$ defined by $F = \sigma\circ X$. In coordinates, if the discrete evolution operator maps $X : (q_0,q_1)\mapsto(q_0,q_1,q_0',q_1')$, then the discrete map will be $F : (q_0,q_1)\mapsto(q_0',q_1')$.

We will be mainly interested in discrete evolution operators which are ***second-order***, which is the requirement that $X(Q\times Q)\subset\ddot{Q}_d$. This implies that they have the form $X : (q_0,q_1)\mapsto(q_0,q_1,q_1,q_2)$, and so the corresponding discrete map will be $F : (q_0,q_1)\mapsto(q_1,q_2)$. We now consider the particular case of a discrete Lagrangian system.

The ***discrete Lagrangian evolution operator*** X_{L_d} is a second-order discrete evolution oper-

ator satisfying

$$D_{\mathrm{DEL}} L_d \circ X_{L_d} = 0$$

and the ***discrete Lagrangian map*** $F_{L_d} : Q \times Q \to Q \times Q$ is defined by $F_{L_d} = \sigma \circ X_{L_d}$.

As in the continuous case, the discrete Lagrangian evolution operator and discrete Lagrangian map are not well defined for arbitrary choices of discrete Lagrangian. We will henceforth assume that L_d is chosen so as to make these structures well defined, and in §2.4 we will give a condition on L_d which ensures that this is true.

A discrete path $q_d \in \mathcal{C}_d(Q)$ is said to be a ***solution of the discrete Euler-Lagrange equations*** if the first term on the right-hand side of (2.9) vanishes for all variations $\delta q_d \in T_{q_d}\mathcal{C}_d(Q)$. This means that the points $\{q_k\}$ satisfy $F_{L_d}(q_{k-1}, q_k) = (q_k, q_{k+1})$ or, equivalently, that they satisfy the ***discrete Euler-Lagrange equations***

$$D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}) = 0, \quad \text{for all } k = 1, \ldots, N-1. \tag{2.11}$$

2.2.2 Discrete Lagrangian maps are symplectic

Define the ***discrete solution space*** $\mathcal{C}_{L_d}(Q) \subset \mathcal{C}_d(Q)$ to be the set of solutions of the discrete Euler-Lagrange equations. Since an element $q_d \in \mathcal{C}_{L_d}(Q)$ is formed by iteration of the map F_{L_d}, it is uniquely determined by the initial condition $(q_0, q_1) \in Q \times Q$. We can thus identify $\mathcal{C}_{L_d}(Q)$ with the space of initial conditions $Q \times Q$.

Defining the ***restricted discrete action map*** $\hat{\mathfrak{G}}_d : Q \times Q \to \mathbb{R}$ to be

$$\hat{\mathfrak{G}}_d(q_0, q_1) = \mathfrak{G}_d(q_d); \quad q_d \in \mathcal{C}_{L_d}(Q) \text{ and } (q_d(t_0), q_d(t_1)) = (q_0, q_1),$$

we see that (2.9) reduces to

$$\begin{aligned} \mathbf{d}\hat{\mathfrak{G}}_d(v_d) \cdot w_{v_d} &= \Theta^+_{L_d}(F^{N-1}_{L_d}(v_d))((F^{N-1}_{L_d})_*(w_{v_d})) - \Theta^-_{L_d}(v_d)(w_{v_d}) \\ &= ((F^{N-1}_{L_d})^*(\Theta^+_{L_d}))(v_d)(w_{v_d}) - \Theta^-_{L_d}(v_d)(w_{v_d}) \end{aligned} \tag{2.12}$$

for all $w_{v_d} \in T_{v_d}(Q \times Q)$ and $v_d = (q_0, q_1) \in Q \times Q$. Taking a further derivative of this expression, and using the fact that $\mathbf{d}^2\hat{\mathfrak{G}}_d = 0$, we obtain

$$(F^{N-1}_{L_d})^*(\Omega_{L_d}) = \Omega_{L_d},$$

where $\Omega_{L_d} = \mathbf{d}\Theta^+_{L_d} = \mathbf{d}\Theta^-_{L_d}$ is the ***discrete Lagrangian symplectic form***, with coordinate expression

$$\Omega_{L_d}(q_0, q_1) = \frac{\partial^2 L_d}{\partial q_0^i \partial q_1^j} \mathbf{d}q_0^i \wedge \mathbf{d}q_1^j.$$

This argument also holds if we take any subinterval of $0, \ldots, N$ and so the statement is true for any number of steps of F_{L_d}. For a single step we have $(F_{L_d})^*\Omega_{L_d} = \Omega_{L_d}$.

Given a map $f : Q \times Q \to Q \times Q$, we will say that f is ***discretely symplectic*** if $f^*\Omega_{L_d} = \Omega_{L_d}$. The above calculations thus prove that the discrete Lagrangian map F_{L_d} is discretely symplectic, just as we saw in the last section that the Lagrangian flow map is symplectic on TQ.

2.2.3 Discrete Noether's theorem

Consider the (left or right) action $\Phi : G \times Q \to Q$ of a Lie group G on Q, with infinitesimal generator as defined in §2.1. This action can be lifted to $Q \times Q$ by the product $\Phi_g^{Q\times Q}(q_0, q_1) = (\Phi_g(q_0), \Phi_g(q_1))$, which has an ***infinitesimal generator*** $\xi_{Q\times Q} : Q \times Q \to T(Q \times Q)$ given by

$$\xi_{Q\times Q}(q_0, q_1) = (\xi_Q(q_0), \xi_Q(q_1)).$$

The two ***discrete Lagrangian momentum maps*** $J^+_{L_d}, J^-_{L_d} : Q \times Q \to \mathfrak{g}^*$ are

$$\begin{aligned} J^+_{L_d}(q_0, q_1) \cdot \xi &= \Theta^+_{L_d} \cdot \xi_{Q\times Q}(q_0, q_1), \\ J^-_{L_d}(q_0, q_1) \cdot \xi &= \Theta^-_{L_d} \cdot \xi_{Q\times Q}(q_0, q_1). \end{aligned}$$

Using the expressions for $\Theta^\pm_{L_d}$ allows the discrete momentum maps to be alternatively written as

$$\begin{aligned} J^+_{L_d}(q_0, q_1) \cdot \xi &= \langle D_2 L_d(q_0, q_1), \xi_Q(q_1) \rangle , \\ J^-_{L_d}(q_0, q_1) \cdot \xi &= \langle -D_1 L_d(q_0, q_1), \xi_Q(q_0) \rangle , \end{aligned}$$

which are computationally useful formulations.

As in the continuous case, it is interesting to consider when the discrete momentum maps are equivariant. This is the conditions

$$\begin{aligned} J^+_{L_d} \circ \Phi_g^{Q\times Q} &= \mathrm{Ad}^*_{g^{-1}} \circ J^+_{L_d}, \\ J^-_{L_d} \circ \Phi_g^{Q\times Q} &= \mathrm{Ad}^*_{g^{-1}} \circ J^-_{L_d}. \end{aligned}$$

In general these equations will not be satisfied; however, there is a simple sufficient condition, similar to the condition in the continuous case.

Recall that we have defined a map $f : Q \times Q \to Q \times Q$ to be discretely symplectic if $f^*\Omega_{L_d} = \Omega_{L_d}$. We now define f to be a ***special discrete symplectic map*** if $f^*\Theta^+_{L_d} = \Theta^+_{L_d}$ and $f^*\Theta^-_{L_d} = \Theta^-_{L_d}$. This clearly means that f is also discretely symplectic, but the reverse is not true.

Theorem 2.5. *Take a discrete Lagrangian system $L_d : Q \times Q \to \mathbb{R}$ with a (left or right) group*

action $\Phi : G \times Q \to Q$. If the product lifted action $\Phi^{Q\times Q} : G \times Q \times Q \to Q \times Q$ acts by special discrete symplectic maps, then the discrete Lagrangian momentum maps are equivariant.

Proof. The proof used in Theorem 2.2 for the continuous case can also be used here, with $J^+_{L_d}$ and $J^-_{L_d}$ being considered separately. □

If the lifted action only preserves one of $\Theta^+_{L_d}$ or $\Theta^-_{L_d}$, then only the corresponding momentum map will necessarily be equivariant.[1]

If a discrete Lagrangian $L_d : Q \times Q \to \mathbb{R}$ is such that $L_d \circ \Phi_g^{Q\times Q} = L_d$ for all $g \in G$, then L_d is said to be ***invariant*** under the lifted action, and Φ is said to be a ***symmetry*** of the discrete Lagrangian. Note that invariance implies ***infinitesimal invariance***, which is $\mathbf{d}L_d \cdot \xi_{Q\times Q} = 0$ for all $\xi \in \mathfrak{g}$. Also note that

$$\mathbf{d}L_d = \Theta^+_{L_d} - \Theta^-_{L_d},$$

and so when L_d is infinitesimally invariant under the lifted action the two discrete momentum maps are equal. In such cases we will use the notation $J_{L_d} : Q \times Q \to \mathfrak{g}^*$ for the unique single ***discrete Lagrangian momentum map***.

Note that invariance of L_d under the lifted action implies that $\Phi_g^{Q\times Q}$ is a special discrete symplectic map. This can be seen by differentiating $L_d \circ \Phi_g^{Q\times Q} = L_d$ with respect to q_1 to obtain

$$D_2L_d\big(\Phi_g^{Q\times Q}(q_0, q_1)\big) \cdot \partial_q\Phi_g(q_1) \cdot \delta q_1 = D_2L_d(q_0, q_1) \cdot \delta q_1,$$

and observing that the left- and right-hand sides are just $(\Phi_g^{Q\times Q})^*\Theta^+_{L_d}$ and $\Theta^+_{L_d}$, respectively, applied to $(q_0, q_1, \delta q_0, \delta q_1)$. Hence $(\Phi_g^{Q\times Q})^*\Theta^+_{L_d} = \Theta^+_{L_d}$, and a similar calculation gives the result for $\Theta^-_{L_d}$.

We now give the discrete analogue of Noether's theorem, Theorem 2.3, which states that momentum maps of symmetries are constants of the motion.

Theorem 2.6 (Discrete Noether's theorem). *Consider a given discrete Lagrangian system $L_d : Q \times Q \to \mathbb{R}$ which is invariant under the lift of the (left or right) action $\Phi : G \times Q \to Q$. Then the corresponding discrete Lagrangian momentum map $J_{L_d} : Q \times Q \to \mathfrak{g}^*$ is a conserved quantity of the discrete Lagrangian map $F_{L_d} : Q \times Q \to Q \times Q$, so that $J_{L_d} \circ F_{L_d} = J_{L_d}$.*

Proof. We will use the same idea as in the proof of the continuous Noether's theorem, based on the fact that the variational principle is invariant under the symmetry action.

Begin by inducing an action of G on the discrete path space $\mathcal{C}_d(Q)$ by using the pointwise action. Then

$$\mathbf{d}\mathfrak{G}_d(q_d) \cdot \xi_{\mathcal{C}_d(Q)}(q_d) = \sum_{k=0}^{N-1} \mathbf{d}L_d \cdot \xi_{Q\times Q},$$

[1] As in the continuous case, equivariance plays an important role in reduction theory and, in the Hamiltonian context, equivariance guarantees that the momentum map is Poisson, which is often useful.

and so the space of solutions $\mathcal{C}_{L_d}(Q)$ of the discrete Euler-Lagrange equations is invariant under the lifted action of G, and the discrete Lagrangian map $F_{L_d} : Q \times Q \to Q \times Q$ commutes with the lifted action $\Phi_g : Q \times Q \to Q \times Q$.

Identifying $\mathcal{C}_{L_d}(Q)$ with the space of initial conditions $Q \times Q$ and using equation (2.12) gives

$$\begin{aligned}\mathbf{d}\mathfrak{G}_d(q_d) \cdot \xi_{\mathcal{C}(Q)}(q_d) &= \mathbf{d}\hat{\mathfrak{G}}_d(q_0, q_1) \cdot \xi_{Q\times Q}(q_0, q_1) \\ &= \left(\left(F^N_{L_d}\right)^* \left(\Theta^+_{L_d}\right) - \Theta^-_{L_d}\right)(q_0, q_1) \cdot \xi_{Q\times Q}(q_0, q_1).\end{aligned}$$

For symmetries the left-hand side is zero, and so we have

$$(\Theta^+_{L_d} \cdot \xi_{Q\times Q}) \circ F^N_{L_d} = \Theta^-_{L_d} \cdot \xi_{Q\times Q},$$

which is simply the statement of preservation of the discrete momentum map, given that for symmetry actions there is only a single momentum map and that the above argument holds for all subintervals, including a single time step. □

As in the continuous case, only infinitesimal invariance of the discrete Lagrangian is actually required for the discrete momentum maps to be conserved. This is due to the fact that only local invariance is used in the proof above, and global invariance is not necessary.

Note that if G is not a symmetry of L_d, then the two discrete momentum maps will not be equal, and it is precisely the difference $J^+_{L_d} - J^-_{L_d}$ which describes the evolution of either momentum map during the time step. To see this, define

$$J^\Delta_{L_d}(q_k, q_{k+1}) = J^+_{L_d}(q_k, q_{k+1}) - J^-_{L_d}(q_k, q_{k+1})$$

and observe that the discrete Euler-Lagrange equations imply

$$J^+_{L_d}(q_{k-1}, q_k) = J^-_{L_d}(q_k, q_{k+1}).$$

Combining the two above expressions shows that the two discrete momentum maps evolve according to

$$\begin{aligned}J^+_{L_d}(q_k, q_{k+1}) &= J^+_{L_d}(q_{k-1}, q_k) + J^\Delta_{L_d}(q_k, q_{k+1}), \\ J^-_{L_d}(q_k, q_{k+1}) &= J^-_{L_d}(q_{k-1}, q_k) + J^\Delta_{L_d}(q_{k-1}, q_k).\end{aligned}$$

Clearly, if L_d is invariant, then $J^\Delta_{L_d} = 0$, and so the momentum maps are equal and they are conserved. If not, then these equations describe how the momentum maps evolve.

2.3 Background: Hamiltonian mechanics

2.3.1 Hamiltonian mechanics

We will only concern ourselves here with the case of a phase space that is the cotangent bundle of a configuration manifold. Although some of the elegance and power of the Hamiltonian formalism is lost in this restriction, it is simpler for our purposes, and of course is the most important case for applications.

Consider then a configuration manifold Q, and define the ***phase space*** to be the cotangent bundle T^*Q. The ***Hamiltonian*** is a function $H : T^*Q \to \mathbb{R}$. We will take local coordinates on T^*Q to be (q, p).

Define the ***canonical one-form*** Θ on T^*Q by

$$\Theta(p_q) \cdot u_{p_q} = \left\langle p_q, T\pi_{T^*Q} \cdot u_{p_q} \right\rangle, \tag{2.13}$$

where $\pi_{T^*Q} : T^*Q \to Q$ is the standard projection and $\langle \cdot, \cdot \rangle$ denotes the natural pairing between vectors and covectors. In coordinates, $\Theta(q, p) = p_i \mathbf{d}q^i$. The ***canonical two-form*** Ω on T^*Q is defined to be

$$\Omega = -\mathbf{d}\Theta,$$

which has coordinate expression $\Omega(q, p) = \mathbf{d}q^i \wedge \mathbf{d}p^i$. The pair (T^*Q, Ω) is an example of a ***symplectic manifold*** and a mapping $F : T^*Q \to T^*Q$ is said to be ***canonical*** or ***symplectic*** if $F^*\Omega = \Omega$. If $F^*\Theta = \Theta$, then F is said to be a ***special symplectic map***, which clearly implies that it is also symplectic. Note that a particular case of special symplectic maps is given by cotangent lifts of maps $Q \to Q$, which automatically preserve the canonical one-form on T^*Q (see Marsden and Ratiu [1999] for details).

Given a Hamiltonian H, define the corresponding ***Hamiltonian vector field*** X_H to be the unique vector field on T^*Q satisfying

$$\mathbf{i}_{X_H}\Omega = \mathbf{d}H. \tag{2.14}$$

Writing $X_H = (X_q, X_p)$ in coordinates, we see that the above expression is

$$-X_{p_i}\mathbf{d}q^i + X_{q^i}\mathbf{d}p_i = \frac{\partial H}{\partial q^i}\mathbf{d}q^i + \frac{\partial H}{\partial p_i}\mathbf{d}p_i,$$

which gives the familiar ***Hamilton's equations*** for the components of X_H, namely,

$$X_{q^i}(q, p) = \frac{\partial H}{\partial p_i}(q, p), \tag{2.15a}$$

$$X_{p_i}(q, p) = -\frac{\partial H}{\partial q^i}(q, p). \tag{2.15b}$$

The ***Hamiltonian flow*** $F_H : T^*Q \times \mathbb{R} \to T^*Q$ is the flow of the Hamiltonian vector field X_H. Note that, unlike the Lagrangian situation, the Hamiltonian vector field X_H and flow map F_H are always well defined for any Hamiltonian.

For any fixed $t \in \mathbb{R}$, the flow map $F_H^t : T^*Q \to T^*Q$ is symplectic, as can be seen by differentiating to obtain

$$\begin{aligned}\left.\frac{\partial}{\partial t}\right|_{t=0} (F_H^t)^*\Omega = \mathcal{L}_{X_H}\Omega &= \mathbf{d}\mathbf{i}_{X_H}\Omega + \mathbf{i}_{X_H}\mathbf{d}\Omega \\ &= \mathbf{d}^2 H - \mathbf{i}_{X_H}\mathbf{d}^2\Theta = 0,\end{aligned}$$

where we have used Cartan's magic formula $\mathcal{L}_X\alpha = \mathbf{d}\mathbf{i}_X\alpha + \mathbf{i}_X\mathbf{d}\alpha$ for the Lie derivative and the fact that $\mathbf{d}^2 = 0$.

2.3.2 Hamiltonian form of Noether's theorem

Consider a (left or right) action $\Phi : G \times Q \to Q$ of G on Q, as in §2.1. The cotangent lift of this action is $\Phi^{T^*Q} : G \times T^*Q \to T^*Q$ given by $\Phi_g^{T^*Q}(p_q) = \Phi_{g^{-1}}^*(p_q)$, which in coordinates is

$$\Phi^{T^*Q}\big(g,(q,p)\big) = \left((\Phi_g^{-1})^i(q), p_j \frac{\partial \Phi_g^j}{\partial q^i}(q)\right).$$

This has the corresponding ***infinitesimal generator*** $\xi_{T^*Q} : T^*Q \to T(T^*Q)$ defined by

$$\xi_{T^*Q}(p_q) = \frac{d}{dg}\big(\Phi_g^{T^*Q}(p_q)\big)\cdot\xi,$$

which has coordinate form

$$\xi_{T^*Q}(q,p) = \left(q^i, p_i, -\left[\left(\frac{\partial\Phi}{\partial q}\right)^{-1}\right]_j^i \frac{\partial\Phi^j}{\partial g^m}\xi^m, \right.$$
$$\left. p_j\frac{\partial^2\Phi^j}{\partial q^i\partial g^m}\xi^m - p_j\frac{\partial^2\Phi_j}{\partial q^i\partial q^j}\left[\left(\frac{\partial\Phi}{\partial q}\right)^{-1}\right]_k^j \frac{\partial\Phi^k}{\partial g^m}\xi^m\right),$$

where the derivatives of Φ are all evaluated at (e,q).

The ***Hamiltonian momentum map*** $J_H : T^*Q \to \mathfrak{g}^*$ is defined by

$$J_H(p_q)\cdot\xi = \Theta(p_q)\cdot\xi_{T^*Q}(p_q).$$

For each $\xi \in \mathfrak{g}$ we define $J_H^\xi : T^*Q \to \mathbb{R}$ by $J_H^\xi(p_q) = J_H(p_q)\cdot\xi$, which has the expression $J_H^\xi = \mathbf{i}_{\xi_{T^*Q}}\Theta$. Note that the Hamiltonian map is also given by the expression

$$J_H(p_q)\cdot\xi = \langle p_q, \xi_Q(q)\rangle,$$

which is useful for computing it in applications.

Writing the requirement for equivariance of a Hamiltonian momentum map gives the equation

$$J_H \circ \Phi_g^{T^*Q} = \mathrm{Ad}^*_{g^{-1}} \circ J_H.$$

Unlike the Lagrangian setting, however, cotangent lifted actions are *always* special symplectic maps, and so we have $(\Phi_g^{T^*Q})^*\Theta = \Theta$ irrespective of the Hamiltonian. This gives the following result.

Theorem 2.7. *Consider a Hamiltonian system $H : T^*Q \to \mathbb{R}$ with a (left or right) group action $\Phi : G \times Q \to Q$. Then the Hamiltonian momentum map $J_H : T^*Q \to \mathfrak{g}^*$ is always equivariant with respect to the cotangent lifted action $\Phi^{T^*Q} : G \times T^*Q \to T^*Q$.*

Proof. Once again, we can use exactly the same proof as for Theorem 2.2 in the continuous case. The only difference is that H need not be restricted to ensure that the lifted action is a special symplectic map. □

A Hamiltonian $H : T^*Q \to \mathbb{R}$ is said to be ***invariant*** under the cotangent lift of the action $\Phi : G \times Q \to Q$ if $H \circ \Phi_g^{T^*Q} = H$ for all $g \in G$, in which case the action is said to be a ***symmetry*** for the Hamiltonian. The derivative of this expression implies that such a Hamiltonian is also ***infinitesimally invariant***, which is the requirement $\mathbf{d}H \cdot \xi_{T^*Q} = 0$ for all $\xi \in \mathfrak{g}$, although the converse is not generally true.

Theorem 2.8 (Hamiltonian Noether's theorem). *Let $H : T^*Q \to \mathbb{R}$ be a Hamiltonian which is invariant under the lift of the (left or right) action $\Phi : G \times Q \to Q$. Then the corresponding Hamiltonian momentum map $J_H : T^*Q \to \mathfrak{g}^*$ is a conserved quantity of the flow; that is, $J_H \circ F_H^t = J_H$ for all times t.*

Proof. Recall that $(\Phi_g^{T^*Q})^*\Theta = \Theta$ for all $g \in G$ as the action is a cotangent lift, and hence $\mathcal{L}_{\xi_{T^*Q}}\Theta = 0$. Now computing the derivative of J_H^ξ in the direction given by the Hamiltonian vector field X_H gives

$$\begin{aligned}
\mathbf{d}J_H^\xi \cdot X_H &= \mathbf{d}(\mathbf{i}_{\xi_{T^*Q}}\Theta) \cdot X_H \\
&= \mathcal{L}_{\xi_{T^*Q}}\Theta \cdot X_H - \mathbf{i}_{\xi_{T^*Q}}\mathbf{d}\Theta \cdot X_H \\
&= -\mathbf{i}_{X_H}\Omega \cdot \xi_{T^*Q} \\
&= -\mathbf{d}H \cdot \xi_{T^*Q} = 0
\end{aligned}$$

using Cartan's magic formula $\mathcal{L}_X\alpha = \mathbf{d}\mathbf{i}_X\alpha + \mathbf{i}_X\mathbf{d}\alpha$ and (2.14). As F_H^t is the flow map for X_H this gives the desired result. □

Noether's theorem still holds even if the Hamiltonian is only infinitesimally invariant, as it is only this local statement which is used in the proof.

2.3.3 Legendre transforms

To relate Lagrangian mechanics to Hamiltonian mechanics we define the ***Legendre transform*** or ***fibre derivative*** $\mathbb{F}L : TQ \to T^*Q$ by

$$\mathbb{F}L(v_q) \cdot w_q = \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} L(v_q + \epsilon w_q),$$

which has coordinate form

$$\mathbb{F}L : (q, \dot{q}) \mapsto (q, p) = \left(q, \frac{\partial L}{\partial \dot{q}}(q, \dot{q})\right).$$

If the fibre derivative of L is locally an isomorphism, then we say that L is ***regular***, and if it is a global isomorphism, then L is said to be ***hyperregular***. We will generally assume that we are working with hyperregular Lagrangians.

The ***fibre derivative*** of a Hamiltonian is the map $\mathbb{F}H : T^*Q \to TQ$ defined by

$$\alpha_q \cdot \mathbb{F}H(\beta_q) = \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} H(\beta_q + \epsilon\alpha_q),$$

which in coordinates is

$$\mathbb{F}H : (q, p) \mapsto (q, \dot{q}) = \left(q, \frac{\partial H}{\partial p}(q, p)\right).$$

Similarly to the situations for Lagrangians, we say that H is ***regular*** if $\mathbb{F}H$ is a local isomorphism, and that H is ***hyperregular*** if $\mathbb{F}H$ is a global isomorphism.

The canonical one- and two-forms and the Hamiltonian momentum maps are related to the Lagrangian one- and two-forms and the Lagrangian momentum maps by pullback under the fibre derivative, so that

$$\Theta_L = (\mathbb{F}L)^*\Theta, \quad \Omega_L = (\mathbb{F}L)^*\Omega, \quad \text{and} \quad J_L = (\mathbb{F}L)^*J_H.$$

If we additionally relate the Hamiltonian to the Lagrangian by

$$H(q, p) = \mathbb{F}L(q, \dot{q}) \cdot \dot{q} - L(q, \dot{q}), \tag{2.16}$$

where (q, p) and $(q, \dot{q})$ are related by the Legendre transform, then the Hamiltonian and Lagrangian vector fields and their associated flow maps will also be related by pullback to give

$$X_L = (\mathbb{F}L)^*X_H; \quad F_L^t = (\mathbb{F}L)^{-1} \circ F_H^t \circ \mathbb{F}L.$$

In coordinates this means that Hamilton's equations (2.15) are equivalent to the Euler-Lagrange equations (2.11). To see this, we compute the derivatives of (2.16) to give

$$\begin{aligned}
\frac{\partial H}{\partial q}(q,p) &= p \cdot \frac{\partial \dot{q}}{\partial q} - \frac{\partial L}{\partial q}(q,\dot{q}) - \frac{\partial L}{\partial \dot{q}}(q,\dot{q})\frac{\partial \dot{q}}{\partial q} \\
&= \frac{\partial L}{\partial q}(q,\dot{q}) && \text{(2.17a)} \\
&= -\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}(q,\dot{q})\right) \\
&= -\dot{p}, && \text{(2.17b)} \\
\frac{\partial H}{\partial \dot{q}}(q,p) &= \dot{q} + p \cdot \frac{\partial \dot{q}}{\partial p} - \frac{\partial L}{\partial \dot{q}}(q,\dot{q})\frac{\partial \dot{q}}{\partial p} \\
&= \dot{q}, && \text{(2.17c)}
\end{aligned}$$

where $p = \mathbb{F}L(q,\dot{q})$ defines $\dot{q}$ as a function of (q,p).

A similar calculation to the above also shows that if L is hyperregular and H is defined by (2.16), then H will also be hyperregular and the fibre derivatives will satisfy $\mathbb{F}H = (\mathbb{F}L)^{-1}$. The converse statement also holds (see Marsden and Ratiu [1999] for more details).

The above relationship between the Hamiltonian and Lagrangian flows can be summarized by the following commutative diagram, where we recall that the symplectic forms and momentum maps are also preserved under each map.

$$\begin{array}{ccc}
TQ & \xrightarrow{F_L^t} & TQ \\
{\scriptstyle \mathbb{F}L}\downarrow & & \downarrow{\scriptstyle \mathbb{F}L} \\
T^*Q & \xrightarrow[F_H^t]{} & T^*Q
\end{array} \tag{2.18}$$

One consequence of this relationship between the Lagrangian and Hamiltonian flow maps is a condition for when the Lagrangian vector field and flow map are well defined.

Theorem 2.9. *Given a Lagrangian $L : TQ \to \mathbb{R}$, the Lagrangian vector field X_L, and hence the Lagrangian flow map F_L, are well defined if and only if the Lagrangian is regular.*

Proof. This can be seen by relating the Hamiltonian and Lagrangian settings with $\mathbb{F}L$, or by computing the Euler-Lagrange equations in coordinates to give

$$\begin{aligned}
0 &= D_1 L(q,\dot{q}) - \frac{d}{dt} D_2 L(q,\dot{q}) \\
&= D_1 L(q,\dot{q}) - D_1 D_2 L(q,\dot{q}) \cdot \dot{q} - D_2 D_2 L(q,\dot{q}) \cdot \ddot{q}.
\end{aligned}$$

Thus, $\ddot{q}$ is well defined as a function of $(q,\dot{q})$ if and only if $D_2 D_2 L$ is invertible, which by the implicit function theorem is equivalent to $\mathbb{F}L$ being locally invertible. □

2.3.4 Generating functions

As with Hamiltonian mechanics, a useful general context for discussing canonical transformations and generating functions is that of symplectic manifolds. Here we limit ourselves, as above, to the case of T^*Q with the canonical symplectic form Ω.

Let $F : T^*Q \to T^*Q$ be a transformation from T^*Q to itself and let $\Gamma(F) \subset T^*Q \times T^*Q$ be the graph of F. Consider the one-form on $T^*Q \times T^*Q$ defined by

$$\hat{\Theta} = \pi_2^*\Theta - \pi_1^*\Theta.$$

where $\pi_i : T^*Q \times T^*Q$ are the projections onto the two components. The corresponding two-form is then

$$\hat{\Omega} = -\mathbf{d}\hat{\Theta} = \pi_2^*\Omega - \pi_1^*\Omega.$$

Denoting the inclusion map by $i_F : \Gamma(F) \to T^*Q \times T^*Q$, we see that we have the identities

$$\pi_1 \circ i_F = \pi_1|_{\Gamma(F)}, \quad \text{and} \quad \pi_2 \circ i_F = F \circ \pi_1 \text{ on } \Gamma(F).$$

Using these relations, we have

$$\begin{aligned} i_F^*\hat{\Omega} &= i_F^*(\pi_2^*\Omega - \pi_1^*\Omega) \\ &= (\pi_2 \circ i_F)^*\Omega - (\pi_1 \circ i_F)^*\Omega \\ &= (\pi_1|_{\Gamma(F)})^*(F^*\Omega - \Omega). \end{aligned}$$

Using this last equality, it is clear that *F is a canonical transformation if and only if* $i_F^*\hat{\Omega} = 0$ or, equivalently, if and only if $\mathbf{d}(i_F^*\hat{\Theta}) = 0$. By the Poincaré lemma, this last statement is equivalent to there existing, at least locally, a function $S : \Gamma(F) \to \mathbb{R}$ such that $i_F^*\hat{\Theta} = \mathbf{d}S$. Such a function S is known as the ***generating function*** of the symplectic transformation F. Note that S is not unique.

The generating function S is specified on the graph $\Gamma(F)$, and so can be expressed in any local coordinate system on $\Gamma(F)$. The standard choices, for coordinates (q_0, p_0, q_1, p_1) on $T^*Q \times T^*Q$, are any two of the four quantities q_0, p_0, q_1 and p_1; note that $\Gamma(F)$ has the same dimension as T^*Q.

2.3.5 Coordinate expression

We will be particularly interested in the choice (q_0, q_1) as local coordinates on $\Gamma(F)$, and so we give the coordinate expressions for the above general generating function derivation for this particular case. This choice results in generating functions of the so-called ***first kind*** [Goldstein, 1980].

Consider a function $S : Q \times Q \to \mathbb{R}$. Its differential is

$$\mathbf{d}S = \frac{\partial S}{\partial q_0}\mathbf{d}q_0 + \frac{\partial S}{\partial q_1}\mathbf{d}q_1.$$

Let $F : T^*Q \to T^*Q$ be the canonical transformation generated by S. In coordinates, the quantity $i_F^*\hat{\Theta}$ is

$$i_F^*\hat{\Theta} = -p_0\mathbf{d}q_0 + p_1\mathbf{d}q_1,$$

and so the condition $i_F^*\hat{\Theta} = \mathbf{d}S$ reduces to the equations

$$p_0 = -\frac{\partial S}{\partial q_0}(q_0, q_1), \tag{2.19a}$$

$$p_1 = \frac{\partial S}{\partial q_1}(q_0, q_1), \tag{2.19b}$$

which are an implicit definition of the transformation $F : (q_0, p_0) \mapsto (q_1, p_1)$. From the above general theory, we know that such a transformation is automatically symplectic, and that all symplectic transformations have such a representation, at least locally.

Note that there is not a one-to-one correspondence between symplectic transformations and real-valued functions on $Q \times Q$, because for some functions the above equations either have no solutions or multiple solutions, and so there is no well-defined map $(q_0, p_0) \mapsto (q_1, p_1)$. For example, taking $S(q_0, q_1) = 0$ forces p_0 to be zero, and so there is no corresponding map φ. In addition, one has to be careful about the special case of generating the identity transformation, as was noted in Channell and Scovel [1990] and Ge and Marsden [1988]. As we will see later, this situation is identical to the existence of solutions to the discrete Euler-Lagrange equations, and, as in that case, we will assume for now that we choose generating functions and time steps so that the equations (2.19) do indeed have solutions.

2.4 Discrete variational mechanics: Hamiltonian viewpoint

2.4.1 Discrete Legendre transforms

Just as the standard Legendre transform maps the Lagrangian state space TQ to the Hamiltonian phase space T^*Q, we can define ***discrete Legendre transforms*** or ***discrete fibre derivatives*** $\mathbb{F}^+L_d, \mathbb{F}^-L_d : Q \times Q \to T^*Q$, which map the discrete state space $Q \times Q$ to T^*Q. These are given by

$$\mathbb{F}^+L_d(q_0, q_1) \cdot \delta q_1 = D_2L_d(q_0, q_1) \cdot \delta q_1,$$

$$\mathbb{F}^-L_d(q_0, q_1) \cdot \delta q_0 = -D_1L_d(q_0, q_1) \cdot \delta q_0,$$

which can be written

$$\mathbb{F}^+L_d : (q_0, q_1) \mapsto (q_1, p_1) = (q_1, D_2L_d(q_0, q_1)),$$
$$\mathbb{F}^-L_d : (q_0, q_1) \mapsto (q_0, p_0) = (q_0, -D_1L_d(q_0, q_1)).$$

If both discrete fibre derivatives are locally isomorphisms (for nearby q_0 and q_1), then we say that L_d is ***regular***. We will generally assume that we are working with regular discrete Lagrangians. In some special cases, such as if Q is a vector space, it may be that both discrete fibre derivatives are global isomorphisms. In that case we say that L_d is ***hyperregular***.

Using the discrete fibre derivatives it can be seen that the canonical one- and two-forms and Hamiltonian momentum maps are related to the discrete Lagrangian forms and discrete momentum maps by pullback, so that

$$\Theta^{\pm}_{L_d} = (\mathbb{F}^{\pm}L_d)^*\Theta, \quad \Omega_{L_d} = (\mathbb{F}^{\pm}L_d)^*\Omega, \quad \text{and} \quad J^{\pm}_{L_d} = (\mathbb{F}^{\pm}L_d)^*J_H.$$

When the discrete momentum maps arise from a symmetry action, the pullback of the Hamiltonian momentum map by either discrete Legendre transform gives the unique discrete momentum map $J_{L_d} = (\mathbb{F}^{\pm}L_d)^*J_H$.

In the continuous case there is a particular relationship between a Lagrangian and a Hamiltonian so that the corresponding vector fields and flow maps are related by pullback under the Legendre transform. Indeed, we rarely consider pairs of Lagrangian and Hamiltonian systems which are not related in this way. In the discrete case a similar relationship exists, as will be shown in §2.5.

Unlike the continuous case, however, we will generally be interested in discrete Lagrangian systems that do not exactly correspond to a given Hamiltonian system. In this case, the symplectic structures and momentum maps are related by pullback under the discrete Legendre transforms, but the flow maps are not. As we will see later, this is a reflection of the fact that discrete Lagrangian systems can be regarded as symplectic-momentum integrators.

2.4.2 Momentum matching

The discrete fibre derivatives also permit a new interpretation of the discrete Euler-Lagrange equations. To see this, we introduce the notation

$$p^+_{k,k+1} = p^+(q_k, q_{k+1}) = \mathbb{F}^+L_d(q_k, q_{k+1}),$$
$$p^-_{k,k+1} = p^-(q_k, q_{k+1}) = \mathbb{F}^-L_d(q_k, q_{k+1}),$$

for the momentum at the two endpoints of each interval $[k, k+1]$. Now observe that the discrete Euler-Lagrange equations are

$$D_2 L_d(q_{k-1}, q_k) = -D_1 L_d(q_k, q_{k+1}),$$

which can be written as

$$\mathbb{F}^+ L_d(q_{k-1}, q_k) = \mathbb{F}^- L_d(q_k, q_{k+1}), \tag{2.20}$$

or simply

$$p^+_{k-1,k} = p^-_{k,k+1}.$$

That is, the discrete Euler-Lagrange equations are simply enforcing the condition that the momentum at time k should be the same when evaluated from the lower interval $[k-1, k]$ or the upper interval $[k, k+1]$. This means that along a solution curve there is a unique momentum at each time k, which we denote by

$$p_k = p^+_{k-1,k} = p^-_{k,k+1}.$$

A discrete trajectory $\{q_k\}_{k=0}^N$ in Q can thus also be regarded as either a trajectory $\{(q_k, q_{k+1})\}_{k=0}^{N-1}$ in $Q \times Q$ or, equivalently, as a trajectory $\{(q_k, p_k)\}_{k=0}^N$ in T^*Q.

It will be useful to note that (2.20) can be written as

$$\mathbb{F}^+ L_d = \mathbb{F}^- L_d \circ F_{L_d}. \tag{2.21}$$

A consequence of viewing the discrete Euler-Lagrange equations as a matching of momenta is that it gives a condition for when the discrete Lagrangian evolution operator and discrete Lagrangian map are well defined.

Theorem 2.10. *Given a discrete Lagrangian system $L_d : Q \times Q \to \mathbb{R}$, the discrete Lagrangian evolution operator X_{L_d} and the discrete Lagrange map F_{L_d} are well defined if and only if $\mathbb{F}^- L_d$ is locally an isomorphism. The discrete Lagrange map is well defined and invertible if and only if the discrete Lagrangian is regular.*

Proof. Given $(q_0, q_1) \in Q \times Q$, the point $q_2 \in Q$ required to satisfy

$$X_{L_d}(q_0, q_1) = (q_0, q_1, q_1, q_2)$$

is defined by equation (2.20), and so q_2 is uniquely defined as a function of q_0 and q_1 if and only if $\mathbb{F}^- L_d$ is locally an isomorphism. From the definition of F_{L_d} it is well defined if and only if X_{L_d} is.

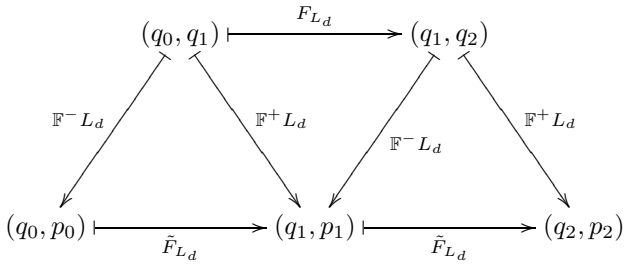
The above argument only implies that F_{L_d} is well defined as a map, however, meaning that it can be applied to map forward in time. For it to be invertible, equation (2.20) shows that it is necessary

and sufficient for $\mathbb{F}^+L_d$ also to be a local isomorphism, which is equivalent to regularity of L_d. □

2.4.3 Discrete Hamiltonian maps

Using the discrete fibre derivatives also enables us to push the discrete Lagrangian map $F_{L_d} : Q \times Q \to Q \times Q$ forward to T^*Q. We define the ***discrete Hamiltonian map*** $\tilde{F}_{L_d} : T^*Q \to T^*Q$ by $\tilde{F}_{L_d} = \mathbb{F}^{\pm}L_d \circ F_{L_d} \circ (\mathbb{F}^{\pm}L_d)^{-1}$. The fact that the discrete Hamiltonian map can be equivalently defined with either discrete Legendre transform is a consequence of the following theorem.

Theorem 2.11. *The following diagram commutes.*



(2.22)

Proof. The central triangle is simply (2.21). Assume that we define the discrete Hamiltonian map by $\tilde{F}_{L_d} = \mathbb{F}^+L_d \circ F_{L_d} \circ (\mathbb{F}^+L_d)^{-1}$, which gives the right-hand parallelogram. Replicating the right-hand triangle on the left-hand side completes the diagram. If we choose to use the other discrete Legendre transform, then the reverse argument applies. □

Corollary 2.1. *The following three definitions of the discrete Hamiltonian map,*

$$\tilde{F}_{L_d} = \mathbb{F}^+L_d \circ F_{L_d} \circ (\mathbb{F}^+L_d)^{-1},$$
$$\tilde{F}_{L_d} = \mathbb{F}^-L_d \circ F_{L_d} \circ (\mathbb{F}^-L_d)^{-1},$$
$$\tilde{F}_{L_d} = \mathbb{F}^+L_d \circ (\mathbb{F}^-L_d)^{-1},$$

are equivalent and have coordinate expression $\tilde{F}_{L_d} : (q_0, p_0) \mapsto (q_1, p_1)$*, where*

$$p_0 = -D_1L_d(q_0, q_1), \tag{2.23a}$$
$$p_1 = D_2L_d(q_0, q_1). \tag{2.23b}$$

Proof. The equivalence of the three definitions can be read directly from the diagram in Theorem 2.11.

The coordinate expression for $\tilde{F}_{L_d} : (q_0, p_0) \mapsto (q_1, p_1)$ can be readily seen from the definition $\tilde{F}_{L_d} = \mathbb{F}^+L_d \circ (\mathbb{F}^-L_d)^{-1}$. Taking initial condition $(q_0, p_0) \in T^*Q$ and setting $(q_0, q_1) = (\mathbb{F}^-L_d)^{-1}(q_0, p_0)$ implies that $p_0 = -D_1L_d(q_0, q_1)$, which is (2.23a). Now, letting $(q_1, p_1) = \mathbb{F}^+L_d(q_0, q_1)$

gives $p_1 = D_2 L_d(q_0, q_1)$, which is (2.23b). □

As the discrete Lagrangian map preserves the discrete symplectic form and discrete momentum maps on $Q \times Q$, the discrete Hamiltonian map will preserve the pushforwards of these structures. As we saw above, however, these are simply the canonical symplectic form and canonical momentum maps on T^*Q, and so the discrete Hamiltonian map is symplectic and momentum-preserving.

We can summarize the relationship between the discrete and continuous systems in the following diagram, where the dashed arrows represent the discretization.

$$\begin{array}{ccc} TQ, F_L & - - \succ & Q \times Q, F_{L_d} \\ \Big\downarrow{\scriptstyle \mathbb{F}L} & & \Big\downarrow{\scriptstyle \mathbb{F}L_d} \\ T^*Q, F_H & - - \succ & T^*Q, \tilde{F}_{L_d} \end{array} \tag{2.24}$$

2.4.4 Discrete Lagrangians are generating functions

As we have seen above, a discrete Lagrangian is a real-valued function on $Q \times Q$ which defines a map $\tilde{F}_{L_d} : T^*Q \to T^*Q$. In fact, a discrete Lagrangian is simply a generating function of the first kind for the map $\tilde{F}_{L_d}$, in the sense defined in §2.3. This is seen by comparing the coordinate expression (2.23) for the discrete Hamiltonian map with the expression (2.19) for the map generated by a generating function of the first kind.

2.5 Correspondence between discrete and continuous mechanics

We will now define a particular choice of discrete Lagrangian which gives an *exact* correspondence between discrete and continuous systems. To do this, we must firstly recall the following fact.

Theorem 2.12. *Consider a regular Lagrangian L for a configuration manifold Q, two points $q_0, q_1 \in Q$ and a time $h \in \mathbb{R}$. If $\|q_1 - q_0\|$ and $|h|$ are sufficiently small, then there exists a unique solution $q : \mathbb{R} \to Q$ of the Euler-Lagrange equations for L satisfying $q(0) = q_0$ and $q(h) = q_1$.*

Proof. See Marsden and Ratiu [1999]. □

For some regular Lagrangian L we now define the ***exact discrete Lagrangian*** to be

$$L_d^E(q_0, q_1, h) = \int_0^h L(q_{0,1}(t), \dot{q}_{0,1}(t))dt \tag{2.25}$$

for sufficiently small h and close q_0 and q_1. Here $q_{0,1}(t)$ is the unique solution of the Euler-Lagrange equations for L which satisfies the boundary conditions $q_{0,1}(0) = q_0$ and $q_{0,1}(h) = q_1$, and whose

existence is guaranteed by Theorem 2.12.

We will now see that with this exact discrete Lagrangian there is an exact correspondence between the discrete and continuous systems. To do this, we will first establish that there is a special relationship between the Legendre transforms of a regular Lagrangian and its corresponding exact discrete Lagrangian. This result will also prove that exact discrete Lagrangians are automatically regular.

Lemma 2.1. *A regular Lagrangian L and the corresponding exact discrete Lagrangian L_d^E have Legendre transforms related by*

$$\begin{aligned}\mathbb{F}^+L_d^E(q_0,q_1,h) &= \mathbb{F}L(q_{0,1}(h),\dot{q}_{0,1}(h)),\\ \mathbb{F}^-L_d^E(q_0,q_1,h) &= \mathbb{F}L(q_{0,1}(0),\dot{q}_{0,1}(0)),\end{aligned}$$

for sufficiently small h and close $q_0, q_1 \in Q$.

Proof. We begin with $\mathbb{F}^-L_d^E$ and compute

$$\begin{aligned}\mathbb{F}^-L_d^E(q_0,q_1,h) &= -\int_0^h \left[\frac{\partial L}{\partial q}\cdot\frac{\partial q_{0,1}}{\partial q_0}+\frac{\partial L}{\partial \dot{q}}\cdot\frac{\partial \dot{q}_{0,1}}{\partial q_0}\right]dt\\ &= -\int_0^h \left[\frac{\partial L}{\partial q}-\frac{d}{dt}\frac{\partial L}{\partial \dot{q}}\right]\cdot\frac{\partial q_{0,1}}{\partial q_0}dt-\left[\frac{\partial L}{\partial \dot{q}}\cdot\frac{\partial q_{0,1}}{\partial q_0}\right]_0^h,\end{aligned}$$

using integration by parts. The fact that $q_{0,1}(t)$ is a solution of the Euler-Lagrange equations shows that the first term is zero. To compute the second term we recall that $q_{0,1}(0)=q_0$ and $q_{0,1}(h)=q_1$, so that

$$\frac{\partial q_{0,1}}{\partial q_0}(0)=\mathrm{Id}\qquad\text{and}\qquad\frac{\partial q_{0,1}}{\partial q_0}(h)=0.$$

Substituting these into the above expression for $\mathbb{F}^-L_d^E$ now gives

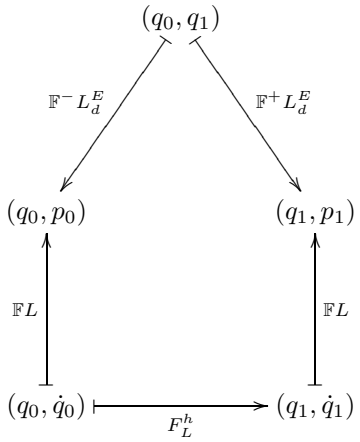
$$\mathbb{F}^-L_d^E(q_0,q_1,h)=\frac{\partial L}{\partial \dot{q}}(q_{0,1}(0),\dot{q}_{0,1}(0)),$$

which is simply the definition of $\mathbb{F}L(q_{0,1}(0),\dot{q}_{0,1}(0))$.

The result for $\mathbb{F}^+L_d^E$ can be established by a similar computation. □

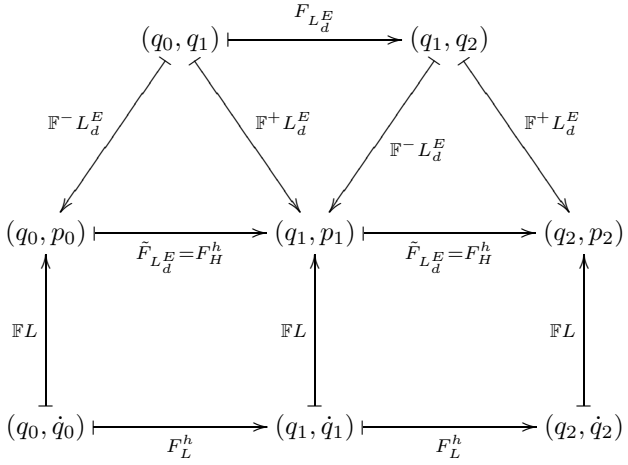
Since $(q_{0,1}(h),\dot{q}_{0,1}(h)) = F_L^h(q_{0,1}(0),\dot{q}_{0,1}(0))$, Lemma 2.1 is equivalent to the following commu-

tative diagram.



(2.26)

Combining this diagram with (2.18) and (2.22) gives the following commutative diagram for the exact discrete Lagrangian.



(2.27)

This proves the following theorem.

Theorem 2.13. *Consider a regular Lagrangian L, its corresponding exact discrete Lagrangian L_d^E, and the pushforward of both the continuous and discrete systems to T^*Q, yielding a Hamiltonian system with Hamiltonian H and a discrete Hamiltonian map $\tilde{F}_{L_d^E}$, respectively. Then, for a sufficiently small time step $h \in \mathbb{R}$, the Hamiltonian flow map equals the pushforward discrete Lagrangian map:*

$$F_H^h = \tilde{F}_{L_d^E}.$$

This theorem is a statement about the time evolution of the system, and can also be interpreted as saying that the diagram (2.24) commutes with the dashed arrows understood as samples at times $\{t_k\}_{k=0}^N$, rather than merely as discretizations.

We can also interpret the equivalence of the discrete and continuous systems as a statement

about trajectories. On the Lagrangian side, this gives the following theorem.

Theorem 2.14. *Take a series of times $\{t_k = kh,\ k = 0, \dots, N\}$ for a sufficiently small time step h, and a regular Lagrangian L and its corresponding exact discrete Lagrangian L_d^E. Then solutions $q : [0, t_N] \to Q$ of the Euler-Lagrange equations for L and solutions $\{q_k\}_{k=0}^N$ of the discrete Euler-Lagrange equations for L_d^E are related by*

$$q_k = q(t_k) \text{ for } k = 0, \dots, N, \tag{2.28a}$$

$$q(t) = q_{k,k+1}(t) \text{ for } t \in [t_k, t_{k+1}]. \tag{2.28b}$$

Here the curves $q_{k,k+1} : [t_k, t_{k+1}] \to Q$ are the unique solutions of the Euler-Lagrange equations for L satisfying $q_{k,k+1}(kh) = q_k$ and $q_{k,k+1}((k+1)h) = q_{k+1}$.

Proof. The main non-obvious issue is smoothness. Let $q(t)$ be a solution of the Euler-Lagrange equations for L and define $\{q_k\}_{k=0}^N$ by (2.28a). Now the discrete Euler-Lagrange equations at time k are simply a matching of discrete Legendre transforms, as in (2.20), but by construction and Lemma 2.1 both sides of this expression are equal to $\mathbb{F}L(q(t_k), \dot{q}(t_k))$. We thus see that $\{q_k\}_{k=0}^N$ is a solution of the discrete Euler-Lagrange equations.

Conversely, let $\{q_k\}_{k=0}^N$ be a solution of the discrete Euler-Lagrange equations for L_d^E and define $q : [0, t_N] \to Q$ by (2.28b). Clearly $q(t)$ is C^2 and a solution of the Euler-Lagrange equations on each open interval (t_k, t_{k+1}), and so we must only establish that it is also C^2 at each t_k, from which it will follow that it is C^2 and a solution on the entire interval $[0, t_N]$.

At time t_k the discrete Euler-Lagrange equations in the form (2.20) together with Lemma 2.1 reduce to

$$\mathbb{F}L(q_{k-1,k}(t_k), \dot{q}_{k-1,k}(t_k)) = \mathbb{F}L(q_{k,k+1}(t_k), \dot{q}_{k,k+1}(t_k)),$$

and, as $\mathbb{F}L$ is a local isomorphism (due to the regularity of L), we see that $q(t)$ is C^1 on $[0, t_N]$. The regularity of L also implies that

$$\ddot{q}(t) = (D_2 D_2 L)^{-1}(D_1 L - D_1 D_2 L \cdot \dot{q}(t))$$

on each open interval (t_k, t_{k+1}), and as the right-hand side only depends on $q(t)$ and $\dot{q}(t)$ this expression is continuous at each t_k, giving that $q(t)$ is indeed C^2 on $[0, t_N]$. □

To summarize, given Lagrangian and Hamiltonian systems with the Legendre transform mapping between them, the symplectic forms and momentum maps are always related by pullback under $\mathbb{F}L$. If, in addition, L and H satisfy the special relationship (2.16), then the flow maps and energy functions will also be related by pullback.

Exactly the same statements hold for the relationship between a discrete Lagrangian system and

a Hamiltonian system. However, when discussing continuous systems we almost always assume that L and H are related by (2.16), whereas for discrete systems we generally do not assume that L_d and L or H are related by (2.25). This is because we are interested in using the discrete mechanics to derive integrators, and the exact discrete Lagrangian is generally not computable.

2.6 Background: Hamilton-Jacobi theory

2.6.1 Generating function for the flow

As discussed in §2.3, it is a standard result that the flow map F_H^t of a Hamiltonian system is a canonical map for each fixed time t. From the generating function theory, it must therefore have a generating function $S(q_0, q_1, t)$. We will now derive a partial differential equation which S must satisfy.

Consider first the time-preserving extension of F_H to the map

$$\hat{F}_H : T^*Q \times \mathbb{R} \to T^*Q \times \mathbb{R}, \quad (p_q, t) \mapsto (F_H^t(p_q), t).$$

Let $\pi_{T^*Q} : T^*Q \times \mathbb{R} \to T^*Q$ be the projection, and define the ***extended canonical one-form*** and the ***extended canonical two-form*** to be

$$\begin{aligned}\Theta_H &= i_{T^*Q}^*\Theta - i_{T^*Q}^*H \wedge \mathbf{d}t,\\ \Omega_H &= -\mathbf{d}\Theta_H = i_{T^*Q}^*\Omega - i_{T^*Q}^*\mathbf{d}H \wedge \mathbf{d}t.\end{aligned}$$

We now calculate

$$\begin{aligned}T\hat{F}_H \cdot \left(\delta p_q, \delta t\right) &= \left(TF_H^t \cdot \delta p_q + \frac{\partial}{\partial t}F_H^t(p_q) \cdot \delta t, \delta t\right)\\ &= (TF_H^t \cdot \delta p_q + X_H \circ F_H^t \cdot \delta t, \delta t),\end{aligned}$$

using that F_H^t is the flow map of the vector field X_H, and so

$$\begin{aligned}\hat{F}_H^*\Omega_H &= (i_{T^*Q} \circ \hat{F}_H)^*\Omega - ((i_{T^*Q} \circ \hat{F}_H)^*\mathbf{d}H) \wedge (\hat{F}_H^*\mathbf{d}t)\\ &= i_{T^*Q}^*(F_H^t)^*\Omega + (i_{T^*Q}^*(F_H^t)^*\mathbf{d}H) \wedge \mathbf{d}t - ((i_{T^*Q} \circ \hat{F}_H)^*\mathbf{d}H) \wedge \mathbf{d}t\\ &= (i_{T^*Q})^*(F_H^t)^*\Omega = i_{T^*Q}^*\Omega\end{aligned}$$

as F_H^t preserves Ω for fixed t. This identity essentially states that the extended flow map pulls back the extended symplectic form to the standard symplectic form.

Consider now the space $T^*Q \times \mathbb{R} \times T^*Q$ and the projection $\pi_1 : T^*Q \times \mathbb{R} \times T^*Q \to T^*Q \times \mathbb{R}$

onto the first two components and $\pi_2 : T^*Q \times \mathbb{R} \times T^*Q \to T^*Q \times \mathbb{R}$ onto the last two components. Define the one-form

$$\hat{\Theta} = \pi_2^* \Theta_H - \pi_1^* i_{T^*Q}^* \Theta,$$

and let the corresponding two-form be

$$\hat{\Omega} = -\mathbf{d}\hat{\Theta} = \pi_2^* \Omega_H - \pi_1^* i_{T^*Q}^* \Omega.$$

The flow map of the Hamiltonian system acts as $F_H : T^*Q \times \mathbb{R} \to T^*Q$ and so the graph of F_H is a subset $\Gamma(F_H) \subset T^*Q \times \mathbb{R} \times T^*Q$. Denote the inclusion map by $i_{F_H} : \Gamma(F_H) \to T^*Q \times \mathbb{R} \times T^*Q$. We now observe that

$$\begin{aligned} \pi_1 \circ i_{F_H} &= \pi_1|_{\Gamma(F_H)}, \\ \pi_2 \circ i_{F_H} &= \hat{F}_H \circ \pi_1 \text{ on } \Gamma(F_H), \end{aligned}$$

and using these relations calculate

$$\begin{aligned} i_{F_H}^* \hat{\Omega} &= i_{F_H}^* \pi_2^* \Omega_H - i_{F_H}^* \pi_1^* i_{T^*Q}^* \Omega \\ &= (\pi_2 \circ i_{F_H})^* \Omega_H - (\pi_1 \circ i_{F_H})^* i_{T^*Q}^* \Omega \\ &= (\pi_1|_{\Gamma(F_H)})^* (\hat{F}_H^* \Omega_H - i_{T^*Q}^* \Omega) \\ &= 0. \end{aligned}$$

We have thus established that $\mathbf{d}(i_{F_H}^* \hat{\Theta}) = 0$ and so, by the Poincaré lemma, there must locally exist a function $S : \Gamma(F_H) \to \mathbb{R}$ so that $i_{F_H}^* \hat{\Theta} = \mathbf{d}S$. It is clear that restricting the above derivations to a section with fixed t simply reproduces the earlier derivation of generating functions for symplectic maps, and so the restriction $S^t : \Gamma(F_H^t) \to \mathbb{R}$ is a generating function for the map $F_H^t : T^*Q \to T^*Q$. The additional information contained in the statement $i_{F_H}^* \hat{\Theta} = \mathbf{d}S$ dictates how S depends on t.

2.6.2 Hamilton-Jacobi equation

As for the case of general generating functions discussed in §2.3, we will now choose a particular set of coordinates on ΓF_H and investigate the implications of $i_{F_H}^* \hat{\Theta} = \mathbf{d}S$.

Consistent with our earlier choice, we will take coordinates (q_0, q_1, t) for $\Gamma(F_H)$ and thus regard the generating function as a map $S : Q \times Q \times \mathbb{R} \to \mathbb{R}$. The differential is thus

$$\mathbf{d}S = \frac{\partial S}{\partial q_0} \mathbf{d}q_0 + \frac{\partial S}{\partial q_1} \mathbf{d}q_1 + \frac{\partial S}{\partial t} \mathbf{d}t,$$

and we also get

$$\hat{\Theta} = -p_0 \mathbf{d}q_0 + p_1 \mathbf{d}q_1 - H(q_1, p_1)\mathbf{d}t,$$

so the condition $i^*_{F_H}\hat{\Theta} = \mathbf{d}S$ is

$$\begin{aligned} p_0 &= -\frac{\partial S}{\partial q_0}(q_0, q_1, t), \\ p_1 &= \frac{\partial S}{\partial q_1}(q_0, q_1, t), \\ H\left(q_1, \frac{\partial S}{\partial q_1}(q_0, q_1, t)\right) &= \frac{\partial S}{\partial t}(q_0, q_1, t). \end{aligned}$$

The first two equations are simply the standard relations which implicitly specify the map F^t_H from the generating function S^t. The third equation specifies the time-dependence of S and is known as the ***Hamilton-Jacobi PDE***, and can be regarded as a partial-differential equation to be solved for S.

To fully specify the Hamilton-Jacobi PDE it is necessary also to provide boundary conditions. As it is first-order in t, it is clear that specifying S as a function of q_0 and q_1 at some time t will define the solution in a neighbourhood of that time. This is equivalent to specifying the map generated by S at some time, up to an arbitrary function of t. Taking this to be the flow map for some fixed time, we see that the unique solution of the Hamilton-Jacobi PDE must be the flow map for nearby t.

2.6.3 Jacobi's solution

While it is possible in principle to solve the Hamilton-Jacobi PDE directly for S, it is generally nonlinear and a closed form solution is not normally possible. By 1840, however, Jacobi had realized that the solution is simply the action of the trajectory joining q_0 and q_1 in time t: see Jacobi [1866]. This is known as ***Jacobi's solution***,

$$S(q_0, q_1, t) = \int_0^t L(q_{0,1}(\tau), \dot{q}_{0,1}(\tau))d\tau, \tag{2.29}$$

where $q_{0,1}(t)$ is a solution of the Euler-Lagrange equations for L satisfying the boundary conditions $q(0) = q_0$ and $q(t) = q_1$, and where L and H are related by the Legendre transform (assumed to be regular). This can be proved in the same way as Lemma 2.1.

2.7 Discrete variational mechanics: Hamilton-Jacobi viewpoint

As was discussed in §2.4, a discrete Lagrangian can be regarded as the generating function for the discrete Hamiltonian map $\tilde{F}_{L_d} : T^*Q \to T^*Q$. We then showed in §2.5 that there is a particular choice of discrete Lagrangian, the so-called *exact* discrete Lagrangian, which exactly generates the flow map F_H of the corresponding Hamiltonian system. From the development of Hamilton-Jacobi theory in §2.6, it is clear that this exact discrete Lagrangian must be a solution of the Hamilton-Jacobi equation. In fact, as can be seen by comparing the definitions given in equations (2.25) and (2.29), the exact discrete Lagrangian is precisely Jacobi's solution of the Hamilton-Jacobi equation.

To summarize, discrete Lagrangian mechanics can be regarded as a variational Lagrangian derivation of the standard generating function and Hamilton-Jacobi theory. Discrete Lagrangians generate symplectic transformations, and given a Lagrangian or Hamiltonian system, one can construct the exact discrete Lagrangian which solves the Hamilton-Jacobi equation, and this will then generate the exact flow of the continuous system.

Chapter 3

Variational integrators

3.1 Introduction

We now turn our attention to considering a discrete Lagrangian system as an approximation to a given continuous system. That is, the discrete system is an integrator for the continuous system.

As we have seen, discrete Lagrangian maps preserve the symplectic structure and so, regarded as integrators, they are necessarily symplectic. Furthermore, generating function theory shows that any symplectic integrator for a mechanical system can be regarded as a discrete Lagrangian system, a fact we state here as a theorem.

Theorem 3.1. *If the integrator $F : T^*Q \times \mathbb{R} \to T^*Q$ is symplectic, then there exists*[1] *a discrete Lagrangian L_d whose discrete Hamiltonian map $\tilde{F}_{L_d}$ is F.*

Proof. As shown above in §2.3, any symplectic transformation locally has a corresponding generating function, which is then a discrete Lagrangian for the method, as discussed in §2.4.4. □

In addition, if the discrete Lagrangian inherits the same symmetry groups as the continuous system, then the discrete system will also preserve the corresponding momentum maps. As an integrator, it will thus be a so-called ***symplectic-momentum integrator***.

Just as with continuous mechanics, we have seen that discrete variational mechanics has both a variational (Lagrangian) and a generating function (Hamiltonian) interpretation. These two viewpoints are complementary and both give insight into the behaviour and derivation of useful integrators.

However, the above theorem is not literally used in the construction of variational integrators, but is rather used as the first steps in obtaining inspiration. We will obtain much deeper insight from the variational principle itself and this is, in large part, what sets variational methods apart from standard symplectic methods.

[1] The discrete Lagrangian may exist only locally, as is the case with generating functions, as was discussed in §2.3.

Symplectic integrators have traditionally been approached from a Hamiltonian viewpoint and there is much existing literature treating this topic (see, for example, Hairer, Nørsett, and Wanner [1993], Hairer and Wanner [1996], MacKay [1992] and Sanz-Serna [1992a]). In this thesis, we concentrate on the analysis of symplectic methods from the variational viewpoint, and we reinterpret many standard concepts from ODE integration theory in this light.

It is also important to distinguish the two ways in which we can derive variational or generating function integrators. First, we can attempt to approximately solve the Hamilton-Jacobi PDE for a given system, such as by taking power series expansions of the generating function. This was used in some of the earliest derivations of symplectic integrators (such as De Vogelaére [1956] and Channell and Scovel [1990]). Second, the method we advocate involves trying to approximate the known Jacobi's solution to the Hamilton-Jacobi PDE: that is, we construct discrete Lagrangians that approximate the exact discrete Lagrangian. This approach is powerful not only because of the coherent and unifying underlying theory that reveals the beautiful geometry underlying discrete mechanics, but also because it leads to practical integrators.

In this section we will assume that Q, and thus also TQ and T^*Q, is a finite-dimensional vector space with an inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\| \cdot \|$. In the case that it is not a vector space, we can embed Q within a vector space and use the theory of constrained discrete systems developed below in §4.4 and discussed further in §4.5.2.

A word of caution: we must be careful about imagining that we can simply pick a coordinate chart and apply the vector space methods described below in such a chart. Doing so indiscriminately can lead to coordinate-dependent integrators that can be unattractive theoretically as well as impractical: for instance, using Euler angles for rigid body integrators has the difficulty that we may spend most of our computational time switching coordinate systems. See, for instance, Wisdom, Peale, and Mignard [1984], Leimkuhler and Patrick [1996] and related papers. For some special classes of configuration manifolds, however, such as when Q is a Lie group, there may be particular global coordinate systems that can be used for this purpose.

We will also frequently consider integrators for Lagrangian systems of the form $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M \dot{q} - V(q)$. When dealing with such systems, we will always assume that M is a positive-definite symmetric mass matrix, so that $\mathbb{F}L(q, \dot{q}) = M\dot{q}$ and thus that L is regular.

3.1.1 Implementation of variational integrators

Although the distinction between the discrete Lagrangian map $F_{L_d} : Q \times Q \times \mathbb{R} \to Q \times Q$ and its pushforward $\tilde{F}_{L_d} : T^*Q \times \mathbb{R} \to T^*Q$ is important geometrically, for implementation purposes the two maps are essentially the same. This is because of the observation made in §2.4.2 that the discrete Euler-Lagrange equations that define F_{L_d} can be interpreted as matching of momenta between adjacent intervals.

In other words, given a trajectory $q_0, q_1, q_2, \ldots, q_{k-1}, q_k$ the map $F_{L_d} : Q \times Q \times \mathbb{R} \to Q \times Q$ calculates q_{k+1} according to

$$D_2 L_d(q_{k-1}, q_k, h) = -D_1 L_d(q_k, q_{k+1}, h).$$

If we now take $p_k = D_2 L_d(q_{k-1}, q_k, h)$ for each k, then this equation is simply

$$p_k = -D_1 L_d(q_k, q_{k+1}, h), \tag{3.1}$$

which, together with the next update

$$p_{k+1} = D_2 L_d(q_k, q_{k+1}, h), \tag{3.2}$$

defines the pushforward map $\tilde{F}_{L_d} : T^*Q \times \mathbb{R} \to T^*Q$. Another way to think of this is that the p_k are merely storing the values $D_2 L_d(q_{k-1}, q_k, h)$ from the last step.

For this reason it is typically easier to implement a variational integrator as the single step map $\tilde{F}_{L_d}$, as this also provides a simple method of initialization from initial values $(q_0, p_0) \in T^*Q$. Many discrete Lagrangians have pushforward maps that are simple to implement. For example, $\tilde{F}_{L_d}$ may be explicit, or it may be a Runge-Kutta method or other integrator type with standard implementation techniques.

In the general case when no special form is apparent, however, the equations (3.1) and (3.2) must be solved directly. The update $(q_k, p_k) \mapsto (q_{k+1}, p_{k+1})$ thus involves first solving the implicit equation (3.1) for q_{k+1} and then evaluating the explicit update (3.2) to give p_{k+1}.

To solve the implicit equation (3.1) we must typically use an iterative technique such as Newton's method. This involves computing a first guess $q_{k+1,0}$ for q_{k+1}, such as $q_{k+1,0} = 2q_k - q_{k-1}$, and then computing the sequence of approximations $q_{k+1,n}$, $n = 1, 2, \ldots$ until they converge to the solution value q_{k+1}. For Newton's method, the iteration rule is given by

$$q^i_{k+1,n+1} = q^i_{k+1,n} - A^{ij}\left[p^j_k + \frac{\partial L_d}{\partial q^j_0}(q_0, q_1, h)\right],$$

where A^{ij} is the inverse of the matrix

$$A_{ij} = \frac{\partial^2 L_d}{\partial q^i_0 \partial q^j_1}(q_0, q_1, h).$$

In the case that the Lagrangian has a simple form, such as $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M \dot{q} - V(q)$, then we can use an initial guess based on p_k, such as $q_{k+1,0} = q_k + hM^{-1}p_k$.

While the Newton's method outlined above typically experiences very fast convergence, it is also

expensive to have to recompute A^{ij} at each iteration of the method. For this reason, it is typical to use an approximation to this matrix which can be held constant for all iterations of Newton's method. See Hairer et al. [1993] for details of this approach for Runge-Kutta methods.

3.1.2 Equivalence of integrators

Given two discrete Lagrangians L_d^1 and L_d^2, we would like to know whether the integrators they generate are the same. Here it will be important to distinguish between the discrete Lagrangian maps $Q \times Q \to Q \times Q$ and the discrete Hamiltonian maps $T^*Q \to T^*Q$. We assume that we are dealing with regular discrete Lagrangians, so that the corresponding maps are well defined.

We say that L_d^1 is (strongly) ***equivalent*** to L_d^2 if their discrete *Hamiltonian* maps are equal, so that $\tilde{F}_{L_d^1} = \tilde{F}_{L_d^2}$. Using the expression $\tilde{F}_{L_d^1} = \mathbb{F}^+ L_d^1 \circ (\mathbb{F}^- L_d^1)^{-1}$, we see that if L_d^1 and L_d^2 are equivalent, then their discrete Legendre transforms will be equal. This implies that the difference $L_d^\Delta = L_d^1 - L_d^2$ must be a function of h only. That is, $L_d^\Delta(q_0, q_1, h) = f(h)$ for some function f. This is clearly also a sufficient condition, as well as being necessary.

We define L_d^1 to be ***weakly equivalent*** to L_d^2 if their discrete *Lagrangian* maps $F_{L_d^1}$ and $F_{L_d^2}$ are equal. A sufficient (and presumably necessary) condition for this to be true is that their difference $L_d^\Delta = L_d^1 - L_d^2$ is a ***null discrete Lagrangian***; that is, the discrete Euler-Lagrange equations for L_d^Δ are satisfied by any triplet (q_0, q_1, q_2). This terminology follows that of the continuous case, as in, for example, Oliver and Sivaloganathan [1988].

If L_d^Δ is a null discrete Lagrangian, then $D_2 L_d^\Delta(q_0, q_1, h)$ cannot depend on q_0 and $D_1 L_d^\Delta(q_1, q_2, h)$ cannot depend on q_2. Furthermore, these two derivatives must be the negative of each other for all q_1. We thus have that L_d^Δ is a null discrete Lagrangian if and only if it is of the form $L_d^\Delta(q_0, q_1, h) = f(q_1, h) - f(q_0, h)$ for some function f.

Using the above calculations, it is clear that strong equivalence implies weak equivalence of discrete Lagrangians. For variational integrators, weak equivalence is thus in some sense the more fundamental notion. Intuitively, if two integrators give solutions which have the same positions q_k for all time, but different momenta p_k at each step, then we would like to regard the methods as being essentially the same. This is exactly weak equivalence.

3.2 Background: Error analysis

In this section we consider a numerical method $F : T^*Q \times \mathbb{R} \to T^*Q$ which approximates the flow $F_H : T^*Q \times \mathbb{R} \to T^*Q$ of a given Hamiltonian vector field X_H. Error analysis is concerned with difference between an exact trajectory and a discrete trajectory.[2]

[2]The reader should be cautioned that in many circumstances, such as the integration of chaotic or complex systems, it may make little sense to imagine accurately computing an exact, but highly unstable, individual trajectory. Instead, we often want to accurately compute robust quantities such as statistical measures that are insensitive to modelling

3.2.1 Local error and method order

An integrator F of X_H is said to be of ***order*** r if there exist an open set $U \subset T^*Q$ and constants $C_l > 0$ and $h_l > 0$ so that

$$\|F(q,p,h) - F_H(q,p,h)\| \leq C_l h^{r+1} \tag{3.3}$$

for all $(q,p) \in U$ and $h \leq h_l$. The expression on the left-hand side of this inequality is known as the ***local error***, and if a method has order of at least 1, then it is said to be ***consistent***.

3.2.2 Global error and convergence

Having defined the error after one step, we now consider the error after many steps. The integrator F of X_H is said to be ***convergent of order*** r if there exist an open set $U \subset T^*Q$ and constants $C_g > 0$, $h_g > 0$ and $T_g > 0$ so that

$$\|(F)^N(q,p,h) - F_H(q,p,T)\| \leq C_g h^r,$$

where $h = T/N$, for all $(q,p) \in U$, $h \leq h_g$ and $T \leq T_g$. The expression on the left-hand side is the ***global error*** at time T.

For one-step methods such as we consider here, convergence follows from a local error bound on the method and a Lipschitz bound on X_H.

Theorem 3.2. *Suppose that the integrator F for X_H is of order r on the open set $U \subset T^*Q$ with local error constant C_l, and assume that $L > 0$ is such that*

$$\left\| \frac{\partial X_H}{\partial (q,p)} \right\| \leq L$$

on U. Then the method is consistent on U with global error constant C_g given by

$$C_g = \frac{C_l}{L}\left(e^{LT_g} - 1\right)$$

Proof. See, for example, Hairer et al. [1993]. □

3.2.3 Order calculation

Given an integrator F for X_H, the order can be calculated by expanding both the true flow F_H and the integrator F in a Taylor series in h and then comparing terms. If the terms agree up to order r, then the method will be of order r.

errors and dynamical sensitivities, and we may wish to do so for long times. In such cases, backward error analysis as outlined in §1.6.5 typically provides a better understanding of the error in the computation. Nonetheless, the forward error is also very important, and we focus on this here.

Here we explicitly write the first few terms of the Taylor series for the true flow for a Hamiltonian of the form $H(q,p) = \frac{1}{2}p^T M^{-1} p + V(q)$. The corresponding Hamiltonian vector field X_H is

$$\dot{q} = M^{-1}p, \tag{3.4a}$$

$$\dot{p} = -\nabla V(q), \tag{3.4b}$$

and so the flow $(q(h), p(h)) = F_H(q_0, p_0, h)$ has the expansion

$$q(h) = q_0 + hM^{-1}p_0 - \frac{1}{2}h^2 M^{-1}\nabla V(q_0) + \mathcal{O}(h^3), \tag{3.5a}$$

$$p(h) = p_0 - h\nabla V(q_0) - \frac{1}{2}h^2 \nabla^2 V(q_0) M^{-1} p_0 + \mathcal{O}(h^3). \tag{3.5b}$$

We will see below an example of using this to calculate the order of a simple class of methods.

3.3 Variational error analysis

Rather than considering how closely the trajectory of F matches the exact trajectory given by F_H, we can alternatively consider how closely a discrete Lagrangian matches the ideal discrete Lagrangian given by the action. As we have seen in §2.5, if the discrete Lagrangian is equal to the action, then the corresponding discrete Hamiltonian map $\tilde{F}_{L_d}$ will exactly equal the flow F_H. We now investigate what happens when this is only an approximation.

The approach taken here is to show that when the discrete Lagrangian approximates a continuous Lagrangian, the discrete integrator approximates the continuous flow and thus the classical theory implies that the global discrete trajectory approximates the continuous trajecctory. An alternative approach, described in Müller and Ortiz [2003], is to directly prove the convergence of trajectories from the convergence of the discrete Lagrangian to a continuous Lagrangian, in an appropriate sense.

3.3.1 Local variational order

Recall that the exact discrete Lagrangian (2.25) is defined by

$$L_d^E(q_0, q_1, h) = \int_0^h L(q, \dot{q}) dt,$$

where $q(t)$ is the solution of the Euler-Lagrange equations satisfying $q(0) = q_0$ and $q(h) = q_1$.

We say that a given discrete Lagrangian L_d is of ***order*** r if there exist an open subset $U_v \subset TQ$ with compact closure and constants $C_v > 0$ and $h_v > 0$ so that

$$\|L_d(q(0), q(h), h) - L_d^E(q(0), q(h), h)\| \leq C_v h^{r+1} \tag{3.6}$$

for all solutions $q(t)$ of the Euler-Lagrange equations with initial condition $(q, \dot{q}) \in U_v$ and for all $h \leq h_v$.

3.3.2 Discrete Legendre transform order

The discrete Legendre transforms $\mathbb{F}^+L_d$ and $\mathbb{F}^-L_d$ of a discrete Lagrangian L_d are said to be of ***order*** r if there exists an open subset $U_f \subset T^*Q$ with compact closure and constants $C_f > 0$ and $h_f > 0$, so that

$$\|\mathbb{F}^+L_d(q(0), q(h), h) - \mathbb{F}^+L_d^E(q(0), q(h), h)\| \leq C_f h^{r+1}, \tag{3.7a}$$

$$\|\mathbb{F}^-L_d(q(0), q(h), h) - \mathbb{F}^-L_d^E(q(0), q(h), h)\| \leq C_f h^{r+1}, \tag{3.7b}$$

for all solutions $q(t)$ of the Euler-Lagrange equations with initial condition $(q, \dot{q}) \in U_f$ and for all $h \leq h_f$.

The relationship between the orders of a discrete Lagrangian, its discrete Legendre transforms and its discrete Hamiltonian map is given in the following fundamental theorem.

Theorem 3.3. *Given a regular Lagrangian L and corresponding Hamiltonian H, the following are equivalent for a discrete Lagrangian L_d:*

(1). the discrete Hamiltonian map for L_d is of order r,

(2). the discrete Legendre transforms of L_d are of order r,

(3). L_d is equivalent to a discrete Lagrangian of order r.

Proof. Begin by assuming that L_d is equivalent to a discrete Lagrangian of order r, and we will show that the discrete Legendre transforms are of order r. From §3.1.2 we know that equivalent discrete Lagrangians have the same discrete Legendre transforms, and we may thus assume without loss that L_d is itself of order r. Now note that (3.6) is equivalent to there existing a function $e_v : T^*Q \times \mathbb{R} \to T^*Q$ so that

$$L_d(q(0), q(h), h) = L_d^E(q(0), q(h), h) + h^{r+1} e_v(q(0), q(h), h)$$

with $\|e_v(q(0), q(h), h)\| \leq C_v$ on U_v. Also, from Theorem 2.12 it is clear that we can parametrize the set U_v by either the initial condition $(q, \dot{q})$ or by the endpoints $(q(0), q(h))$.

Taking derivatives of the above expression with respect to $q(h)$ gives

$$\mathbb{F}^+L_d(q(0), q(h), h) = \mathbb{F}^+L_d^E(q(0), q(h), h) + h^{r+1} D_2 e_v(q(0), q(h), h),$$

and as e_v is smooth and bounded on the compact set $\mathrm{cl}(U_v)$, so too is $D_2 e_v$, giving (3.7a). Taking derivatives with respect to $q(0)$ now shows that the discrete Legendre transforms of L_d are of order r.

Now assume that $\mathbb{F}^+ L_d$ and $\mathbb{F}^- L_d$ are of order r, and set

$$e_v(q(0), q(h), h) = \frac{1}{h^{r+1}} \big[L_d(q(0), q(h), h) - L_d^E(q(0), q(h), h) \big].$$

Taking derivatives with respect to $q(0)$ and $q(h)$ and using (3.7) shows that $\|D_1 e_v\| \leq C_f$ and $\|D_2 e_v\| \leq C_f$ on $\mathrm{cl}(U_f)$, which is compact. This then implies that $e_v(\cdot, \cdot, h)$ is itself locally bounded in its first two arguments, and so there exists a function $d(h)$ and a constant C_v such that $\|e_v(q(0), q(h), h) - d(h)\| \leq C_v$. This then proves that $L_d(q(0), q(h), h) - d(h)$ has variational order r, and so L_d is equivalent to a discrete Lagrangian of order r.

We will now show the equivalence of the discrete Hamiltonian map being of order r and the discrete Legendre transforms being of order r. To do this we will make use of the following fact, which is a consequence of the implicit function theorem.

Assume that we have smooth functions related by

$$\begin{aligned} f_1(x, h) &= g_1(x, h) + h^{r+1} e_1(x, h), \\ f_2(y, h) &= g_2(y, h) + h^{r+1} e_2(y, h), \end{aligned}$$

with e_1 and e_2 bounded on some compact sets. Then we have

$$f_2(f_1(x, h), h) = g_2(g_1(x, h), h) + h^{r+1} e_{12}(x, h), \tag{3.8a}$$

$$f_1^{-1}(y, h) = g_1^{-1}(y, h) + h^{r+1} \bar{e}_1(y, h), \tag{3.8b}$$

for some functions $e_{12}(x, h)$ and $\bar{e}_1(y, h)$ bounded on compact sets.

Now assume that $\mathbb{F}^+ L_d$ and $\mathbb{F}^- L_d$ are of order r and use Corollary 2.1 to write

$$\begin{aligned} \tilde{F}_{L_d} &= \mathbb{F}^+ L_d \circ (\mathbb{F}^- L_d)^{-1}, \\ \tilde{F}_{L_d^E} &= \mathbb{F}^+ L_d^E \circ (\mathbb{F}^- L_d^E)^{-1}. \end{aligned}$$

Equation (3.8) gives the existence of a bounded function e_l such that

$$\tilde{F}_{L_d}(q(0), q(h), h) = \tilde{F}_{L_d^E}(q(0), q(h), h) + h^{r+1} e_l(q(0), q(h), h),$$

and thus we see that the discrete Hamiltonian map is of order r.

Finally, assume that $\tilde{F}_{L_d^E}$ is of order r, and observe that

$$(\mathbb{F}^- L_d)^{-1}(q_0, p_0) = (q_0, \pi_Q \circ \tilde{F}_{L_d}(q_0, p_0)),$$

so (3.8) implies (3.7b). But now we recall from (2.21) that

$$\mathbb{F}^+ L_d = \tilde{F}_{L_d} \circ \mathbb{F}^- L_d,$$

and together with (3.8a) this gives (3.7a), showing that the discrete Legendre transforms are of order r. □

3.3.3 Variational order calculation

Given a discrete Lagrangian, its order can be calculated by expanding the expression for $L_d(q(0), q(h), h)$ in a Taylor series in h and comparing this to the same expansion for the exact Lagrangian. If the series agree up to r terms, then the discrete Lagrangian is of order r.

We explicitly evaluate the first few terms of the expansion of the exact discrete Lagrangian to give

$$L_d^E(q(0), q(h), h) = hL(q, \dot{q}) + \frac{1}{2}h^2 \left(\frac{\partial L}{\partial q}(q, \dot{q}) \cdot \dot{q} + \frac{\partial L}{\partial \dot{q}}(q, \dot{q}) \cdot \ddot{q} \right) + \mathcal{O}(h^3), \tag{3.9}$$

where $q = q(0)$, $\dot{q} = \dot{q}(0)$ and so forth. Higher derivatives of $q(t)$ are determined by the Euler-Lagrange equations.

Example 3.1. *An illustrative class of discrete Lagrangian is given by*

$$L_d^\alpha(q_0, q_1, h) = hL\left((1-\alpha)q_0 + \alpha q_1, \frac{q_1 - q_0}{h} \right)$$

for some parameter $\alpha \in [0, 1]$. Calculating the expansion in h gives

$$L_d^\alpha(q(0), q(h), h) = hL(q, \dot{q}) + \frac{1}{2}h^2 \left(2\alpha \frac{\partial L}{\partial q}(q, \dot{q}) \cdot \dot{q} + \frac{\partial L}{\partial \dot{q}}(q, \dot{q}) \cdot \ddot{q} \right) + \mathcal{O}(h^3).$$

Comparing this to the expansion (3.9) for the exact discrete Lagrangian shows that the method is second-order if and only if $\alpha = 1/2$; otherwise it is only consistent.

Calculating the discrete Hamiltonian map for $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M \dot{q} - V(q)$ gives the integrator $\tilde{F}_{L_d}^\alpha : (q_0, p_0) \mapsto (q_1, p_1)$ defined implicitly by the relations

$$\frac{q_1 - q_0}{h} = M^{-1} \left(\alpha p_0 + (1-\alpha) p_1 \right), \tag{3.10a}$$

$$\frac{p_1 - p_0}{h} = -\nabla V \left((1-\alpha) q_0 + \alpha q_1 \right). \tag{3.10b}$$

Note that this method is explicit for $\alpha = 0$ or $\alpha = 1$ and that it is simply the midpoint rule for $\alpha = 1/2$. Expanding (3.10) in h gives

$$q_1 = q_0 + hM^{-1}p_0 - (1-\alpha)h^2\nabla V(q_0) + \mathcal{O}(h^3),$$
$$p_1 = p_0 - h\nabla V(q_0) - \alpha h^2\nabla^2 V(q_0)M^{-1}p_0 + \mathcal{O}(h^3),$$

and comparing this to the expansion (3.5) of the true flow shows that the method is second-order if and only if $\alpha = 1/2$, and otherwise it is only consistent.

The local error and the discrete Lagrangian error thus agree, as expected. ◊

Example 3.2. *As the expansions of discrete Lagrangians are linear in L_d, if we take the symmetrized discrete Lagrangian*

$$L_d^{sym,\alpha} = \frac{1}{2}L_d^{\alpha} + \frac{1}{2}L_d^{1-\alpha},$$

then the expansion will agree with that of the exact discrete Lagrangian up to terms of order h^2, so it gives a method that is second-order for any α. ◊

3.4 The adjoint of a method and symmetric methods

For a one-step method $F : T^*Q \times \mathbb{R} \to T^*Q$ the *adjoint method* is $F^* : T^*Q \times \mathbb{R} \to T^*Q$ defined by

$$(F^*)^h \circ F^{-h} = \mathrm{Id}; \tag{3.11}$$

that is, $(F^*)^h = (F^{-h})^{-1}$. The method F is said to be *self-adjoint* if $F^* = F$. Note that we always have $F^{**} = F$.

Given a discrete Lagrangian $L_d : Q \times Q \times \mathbb{R} \to \mathbb{R}$, we define the *adjoint discrete Lagrangian* to be $L_d^* : Q \times Q \times \mathbb{R} \to \mathbb{R}$ defined by

$$L_d^*(q_0, q_1, h) = -L_d(q_1, q_0, -h). \tag{3.12}$$

The discrete Lagrangian L_d is said to be *self-adjoint* if $L_d^* = L_d$. Note that $L_d^{**} = L_d$ for any L_d.

Theorem 3.4. *If the discrete Lagrangian L_d has discrete Hamiltonian map $\tilde{F}_{L_d}$, then the adjoint L_d^* of the discrete Lagrangian has discrete Hamiltonian map equal to the adjoint map, so that $\tilde{F}_{L_d^*} = \tilde{F}_{L_d}^*$. If the discrete Lagrangian is self-adjoint, then the method is self-adjoint. Conversely, if the method is self-adjoint, then the discrete Lagrangian is equivalent to a self-adjoint discrete Lagrangian.*

Proof. Consider discrete Lagrangians L_d and L_d^* and the corresponding discrete Hamiltonian maps $\tilde{F}_{L_d}$ and $\tilde{F}_{L_d^*}$. For $\tilde{F}_{L_d}$ and $\tilde{F}_{L_d^*}$ to be adjoint, the definition (3.11) requires that $\tilde{F}_{L_d}(q_0, p_0, -h) =$

(q_1, p_1) and $\tilde{F}_{L_d^*}(q_1, p_1, h) = (q_0, p_0)$ for all (q_0, p_0). In terms of the generating functions this is

$$\begin{aligned} p_0 &= -D_1 L_d(q_0, q_1, -h), \\ p_1 &= D_2 L_d(q_0, q_1, -h), \\ p_1 &= -D_1 L_d^*(q_1, q_0, h), \\ p_0 &= D_2 L_d^*(q_1, q_0, h). \end{aligned} \tag{3.13}$$

Equating the expressions for p_0 and p_1 shows that this, in turn, is equivalent to

$$\begin{aligned} -D_1 L_d(q_0, q_1, -h) &= D_2 L_d^*(q_1, q_0, h), \\ D_2 L_d(q_0, q_1, -h) &= D_1 L_d^*(q_1, q_0, h). \end{aligned} \tag{3.14}$$

Now, if L_d and L_d^* are mutually adjoint, then the definition (3.12) implies (3.14) and so (3.13), thus establishing that $\tilde{F}_{L_d}$ and $\tilde{F}_{L_d^*}$ must also be mutually adjoint, which is written $\tilde{F}_{L_d^*} = \tilde{F}_{L_d}^*$. Note that this implies that $\tilde{F}_{L_d^*}^* = \tilde{F}_{L_d}$.

If L_d is self-adjoint and so $L_d = L_d^*$, then this immediately gives that $\tilde{F}_{L_d} = \tilde{F}_{L_d}^*$ and so $\tilde{F}_{L_d}$ is also self-adjoint.

Conversely, if $\tilde{F}_{L_d}$ and $\tilde{F}_{L_d^*}$ are adjoint, then (3.11) implies (3.13) which implies (3.14). As this simply states that the derivatives of L_d and L_d^* with respect to q_0 and q_1 satisfy the requirement (3.12) for adjointness it follows that L_d and L_d^* are mutually adjoint up to the addition of a function of h. Symmetry of $\tilde{F}_{L_d}$ thus implies symmetry of L_d up to a function of h, and so L_d is equivalent to a self-adjoint discrete Lagrangian. □

3.4.1 Exact discrete Lagrangian is self-adjoint

It is easy to verify that the exact discrete Lagrangian (2.25) is self-adjoint. This can be done either directly from the definition (3.12), or by realizing that the exact flow map F_H generated by L_d^E satisfies (3.11), and then using Theorem 3.4.

3.4.2 Order of adjoint methods

To relate the expansions of L_d and its adjoint in terms of h, it is necessary to work with the modified form

$$L_d^*\big(q(-h/2), q(h/2), h\big) = -L_d\big(q(h/2), q(-h/2), -h\big),$$

which can be used in the same way as $L_d^*(q(0), q(h), h) = -L_d(q(h), q(0), -h)$. From this it is clear that the expansion of L_d^* is the negative of the expansion of L_d with h replaced by $-h$. In other

words, if L_d has expansion

$$L_d(h) = hL_d^{(1)} + \frac{1}{2}h^2 L_d^{(2)} + \frac{1}{6}h^3 L_d^{(3)} + \cdots,$$

then L_d^* will have expansion

$$\begin{aligned} L_d^*(h) &= -(-h)L_d^{(1)} - \frac{1}{2}(-h)^2 L_d^{(2)} - \frac{1}{6}(-h)^3 L_d^{(3)} - \cdots \\ &= hL_d^{(1)} - \frac{1}{2}h^2 L_d^{(2)} + \frac{1}{6}h^3 L_d^{(3)} - \cdots \end{aligned}$$

and so the series agree on odd terms and are opposite on even terms.

This shows that the order of the adjoint discrete Lagrangian L_d^* is the same as the order of L_d. Furthermore, if L_d is self-adjoint, then all the even terms in its expansion must be zero, showing that self-adjoint discrete Lagrangians are necessarily of even order (the first nonzero term, which is $r+1$, must be odd).

These same conclusions can be also be reached by working with the discrete Hamiltonian map, and showing that its adjoint has the same order as it, and that it is of even order whenever it is self-adjoint. Theorems 3.4 and 3.3 then give the corresponding statements for the discrete Lagrangians.

Example 3.3. *Perhaps the simplest example of adjoint discrete Lagrangians is the pair*

$$\begin{aligned} L_d(q_0, q_1, h) &= hL\left(q_0, \frac{q_1 - q_0}{h}\right), \\ L_d^*(q_0, q_1, h) &= hL\left(q_1, \frac{q_1 - q_0}{h}\right), \end{aligned}$$

which clearly satisfy (3.12). For a Lagrangian of the form $L = \frac{1}{2}\dot{q}^T M \dot{q} - V(q)$ these two discrete Lagrangians produce the methods $\tilde{F}_{L_d}$ and $\tilde{F}_{L_d^*}$ given by

$$\tilde{F}_{L_d} \begin{cases} q_1 = q_0 + hM^{-1}p_1, \\ p_1 = p_0 - h\nabla V(q_0), \end{cases}$$

$$\tilde{F}_{L_d^*} \begin{cases} q_1 = q_0 + hM^{-1}p_0, \\ p_1 = p_0 - h\nabla V(q_1). \end{cases}$$

In the terminology of Hairer et al. [2002] these are the two types of symplectic Euler. We can now explicitly compute:

$$\begin{aligned} (\tilde{F}_{L_d^*})^h \circ (\tilde{F}_{L_d})^{(-h)}(q_0, p_0) &= \tilde{F}_{L_d^*}(q_0 + hM^{-1}p_1, p_0 - h\nabla V(q_0), h) \\ &= (q_0, p_0), \end{aligned}$$

which shows that $\tilde{F}_{L_d}$ and $\tilde{F}_{L_d^*}$ are indeed mutually adjoint. $\Diamond$

Example 3.4. *The discrete Lagrangians in the previous example are just L_d^α for $\alpha = 0$ and $\alpha = 1$, respectively. Extending this gives $(L_d^\alpha)^* = L_d^{1-\alpha}$, which shows that the midpoint rule (given by $\alpha = 1/2$) is self-adjoint. From this it is also clear that the symmetrized versions $L_d^{sym,\alpha}$ are self-adjoint for all α.* $\Diamond$

3.5 Composition methods

We now consider how to combine several discrete Lagrangians together to obtain a new discrete Lagrangian with higher order, or some other desirable property. The resulting discrete Hamiltonian map will be the composition of the maps of the component discrete Lagrangians. References on composition methods include Yoshida [1990], Qin and Zhu [1992], McLachlan [1993] and Murua and Sanz-Serna [1999].

The strength of the composition methodology can be illustrated by a few simple examples. Given a one-step method $F : T^*Q \times \mathbb{R} \to T^*Q$ with corresponding adjoint F^*, then the method $\hat{F}^h = F^{h/2} \circ (F^*)^{h/2}$ will be self-adjoint and have order at least equal to that of F. Furthermore, for a self-adjoint method F with order r, which we recall must be even, the method $\hat{F}^h = F^{\gamma h} \circ F^{(1-2\gamma)h} \circ F^{\gamma h}$ with the constant $\gamma = (2 - 2^{1/(r+1)})^{-1}$ will have order $r + 2$. This thus provides a simple way to derive methods of arbitrarily high order starting from a given low-order method. See the above references for details and more complicated examples.

Consider now discrete Lagrangians L_d^i and time step fractions γ^i for $i = 1, \ldots, s$ satisfying $\sum_{i=1}^s \gamma^i = 1$. Note that the γ^i may each be positive or negative. We now give three equivalent interpretations of composition discrete Lagrangians.

3.5.1 Multiple steps

Begin by taking a discrete trajectory $\{q_k\}_{k=0}^N$, dividing each step (q_k, q_{k+1}) into s substeps ($q_k = q_k^0, q_k^1, q_k^2, \ldots, q_k^s = q_{k+1}$). Rather than using the same discrete Lagrangian for each step, as we have previously always assumed, we will now use the different L_d^i on each substep in turn.

This is equivalent to taking the discrete action sum to be

$$\mathfrak{G}_d\big(\{(q_k = q_k^0, \ldots, q_k^s = q_{k+1})\}_{k=1}^N\big) = \sum_{k=0}^{N} \sum_{i=1}^{s} L_d^i(q_k^{i-1}, q_k^i, \gamma^i h). \tag{3.15}$$

The discrete Euler-Lagrange equations, resulting from requiring that this be stationary, pair neigh-

bouring discrete Lagrangians together to give

$$D_2 L_d^i(q_k^{i-1}, q_k^i, \gamma^i h) + D_1 L_d^{i+1}(q_k^i, q_k^{i+1}, \gamma^{i+1} h) = 0, \qquad i = 1, \ldots, s-1, \tag{3.16a}$$

$$D_2 L_d^s(q_k^{s-1}, q_k^s, \gamma^s h) + D_1 L_d^1(q_{k+1}^0, q_{k+1}^1, \gamma^1 h) = 0, \tag{3.16b}$$

where the steps are joined with $q_k^s = q_{k+1}^0$.

Considering the L_d^i as generating functions for the discrete Hamiltonian maps $\tilde{F}_{L_d^i}$ shows that this is simply taking a step with $\tilde{F}_{L_d^1}$ of length $\gamma^1 h$, followed by a step with $\tilde{F}_{L_d^2}$ of length $\gamma^2 h$, and so on. The map over the entire time step is thus the composition of the maps

$$\tilde{F}_{L_d^s}^{\gamma^s h} \circ \cdots \circ \tilde{F}_{L_d^2}^{\gamma^2 h} \circ \tilde{F}_{L_d^1}^{\gamma^1 h}.$$

3.5.2 Single step, multiple substeps

We now combine the discrete Lagrangians on each step into one multipoint discrete Lagrangian defined by

$$L_d(q_k^0, q_k^1, \ldots, q_k^s, h) = \sum_{i=1}^{s} L_d^i(q_k^{i-1}, q_k^i, \gamma^i h), \tag{3.17}$$

and we define the discrete action sum over the entire trajectory to be

$$\mathfrak{G}_d\big(\{(q_k = q_k^0, \ldots, q_k^s = q_{k+1})\}_{k=1}^N\big) = \sum_{k=0}^{N} L_d(q_k^0, q_k^1, \ldots, q_k^s, h), \tag{3.18}$$

which is clearly equal to (3.15).

Requiring that $\mathfrak{G}_d$ be stationary gives the extended set of discrete Euler-Lagrange equations

$$D_i L_d(q_k^0, q_k^1, \ldots, q_k^s, h) = 0 \qquad i = 2, \ldots, s \tag{3.19a}$$

$$D_{s+1} L_d(q_k^0, q_k^1, \ldots, q_k^s, h) + D_1 L_d(q_{k+1}^0, q_{k+1}^1, \ldots, q_{k+1}^s, h) = 0, \tag{3.19b}$$

which are equivalent to (3.16a) and (3.16b), respectively.

3.5.3 Single step

Finally, we form a standard discrete Lagrangian which is equivalent to the above methods. Set the ***composition discrete Lagrangian*** to be

$$L_d(q_k, q_{k+1}, h) = \underset{(q_k^1, \ldots, q_k^{s-1})}{\operatorname{ext}} L_d(q_k = q_k^0, q_k^1, q_k^2, \ldots, q_k^{s-1}, q_k^s = q_{k+1}, h) \tag{3.20}$$

which is the multipoint discrete Lagrangian defined above, evaluated on the trajectory within the step which solves (3.19a).

Note that the derivatives of this discrete Lagrangian satisfy

$$\begin{aligned} D_1 L_d(q_k, q_{k+1}, h) &= D_1 L_d(q_k, q_k^1, q_k^2, \ldots, q_k^{s-1}, q_{k+1}, h) \\ &\quad + \sum_{i=1}^{s-1} D_i L_d(q_k, q_k^1, q_k^2, \ldots, q_k^{s-1}, q_{k+1}, h) \cdot \frac{\partial q_k^i}{\partial q_k} \\ &= D_1 L_d(q_k, q_k^1, q_k^2, \ldots, q_k^{s-1}, q_{k+1}, h) \\ &= D_1 L_d^1(q_k, q_k^1, \gamma^1 h) \end{aligned}$$

using (3.19a), and similarly

$$\begin{aligned} D_2 L_d(q_k, q_{k+1}, h) &= D_{s+1} L_d(q_k, q_k^1, q_k^2, \ldots, q_k^{s-1}, q_{k+1}, h) \\ &= D_2 L_d^s(q_k^{s-1}, q_{k+1}, \gamma^s h). \end{aligned}$$

This gives the following theorem.

Theorem 3.5. *Take discrete Lagrangians L_d^i and time step fractions γ^i for $i = 1, \ldots, s$ satisfying $\sum_{i=1}^s \gamma^i = 1$. Define the composition discrete Lagrangian L_d by (3.20). Then the discrete Hamiltonian map $\tilde{F}_{L_d}$ is*

$$\tilde{F}_{L_d}^h = \tilde{F}_{L_d^s}^{\gamma^s h} \circ \cdots \circ \tilde{F}_{L_d^2}^{\gamma^2 h} \circ \tilde{F}_{L_d^1}^{\gamma^1 h}$$

formed by the composition of the discrete Hamiltonian maps for each L_d^i.

Proof. The equations that define $\tilde{F}_{L_d}$ are

$$\begin{aligned} p_k &= -D_1 L_d(q_k, q_{k+1}, h) = -D_1 L_d^1(q_k, q_k^1, \gamma^1 h), \\ p_{k+1} &= D_2 L_d(q_k, q_{k+1}, h) = D_2 L_d^s(q_k^{s-1}, q_{k+1}, \gamma^s h), \end{aligned}$$

together with (3.19a), which is equivalent to (3.16a), which we write as

$$p_k^i = D_2 L_d^i(q_k^{i-1}, q_k^i, \gamma^i h) = -D_1 L_d^{i+1}(q_k^i, q_k^{i+1}, \gamma^{i+1} h)$$

for $i = 1, \ldots, s-1$. Setting $p_k^0 = p_k$ and $p_k^s = p_{k+1}$, we can group these to give

$$\begin{aligned} p_k^{i-1} &= -D_1 L_d^i(q_k^{i-1}, q_k^i, \gamma^i h), \\ p_k^i &= D_2 L_d^i(q_k^{i-1}, q_k^i, \gamma^i h), \end{aligned}$$

for $i = 1, \ldots, s$, which are the definition of $\tilde{F}_{L_d^s}^{\gamma^s h} \circ \cdots \circ \tilde{F}_{L_d^2}^{\gamma^2 h} \circ \tilde{F}_{L_d^1}^{\gamma^1 h}$, thus giving the required

equivalence. □

3.6 Examples of variational integrators

In this section we will consider a number of standard symplectic methods and show how to write them as variational integrators. Recall that we are assuming that Q is a linear space with inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\| \cdot \|$. We will always assume that the Lagrangian $L : TQ \to \mathbb{R}$ is regular, so that it has a corresponding Hamiltonian $H : T^*Q \to \mathbb{R}$. In addition, we will sometimes consider the Lagrangian composed of a kinetic and a potential energy, so that it is of the form $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M \dot{q} - V(q)$, where M is a positive-definite symmetric matrix.

3.6.1 Midpoint rule

Given a Hamiltonian system $H : T^*Q \to \mathbb{R}$, the ***midpoint rule*** is an integrator $F^h : (q_0, p_0) \mapsto (q_1, p_1)$. Setting $z_0 = (q_0, p_0)$ and $z_1 = (q_1, p_1)$ the map is defined implicitly by the relation

$$\frac{z_1 - z_0}{h} = X_H \left(\frac{z_1 + z_0}{2} \right),$$

where X_H is the Hamiltonian vector field. Writing the two components separately gives

$$\frac{q_1 - q_0}{h} = \frac{\partial H}{\partial p} \left(\frac{q_1 + q_0}{2}, \frac{p_1 + p_0}{2} \right), \tag{3.21a}$$

$$\frac{p_1 - p_0}{h} = -\frac{\partial H}{\partial q} \left(\frac{q_1 + q_0}{2}, \frac{p_1 + p_0}{2} \right). \tag{3.21b}$$

The symplectic nature of the midpoint rule is often explained by using the Cayley transform (this remark is due, as far as we know, to Krishnaprasad and J. C. Simo; see, for example, M. A. Austin and Wang [1993], Simo et al. [1992] and Simo and Tarnow [1992], and related papers). See Marsden [1999] for an exposition of this method.

To write the midpoint rule as a variational integrator, assume that H is regular and that L is the corresponding regular Lagrangian defined by (2.16). Define the discrete Lagrangian

$$L_d^{\frac{1}{2}}(q_0, q_1, h) = hL \left(\frac{q_1 + q_0}{2}, \frac{q_1 - q_0}{h} \right).$$

Evaluating the expressions (2.23) for the discrete Hamiltonian map gives

$$p_0 = -\frac{h}{2}\frac{\partial L}{\partial q} \left(\frac{q_1 + q_0}{2}, \frac{q_1 - q_0}{h} \right) + \frac{\partial L}{\partial \dot{q}} \left(\frac{q_1 + q_0}{2}, \frac{q_1 - q_0}{h} \right),$$

$$p_1 = \frac{h}{2}\frac{\partial L}{\partial q} \left(\frac{q_1 + q_0}{2}, \frac{q_1 - q_0}{h} \right) + \frac{\partial L}{\partial \dot{q}} \left(\frac{q_1 + q_0}{2}, \frac{q_1 - q_0}{h} \right),$$

and subtracting and adding these two equations produces

$$\begin{aligned} \frac{p_1 - p_0}{h} &= \frac{\partial L}{\partial q}\left(\frac{q_1+q_0}{2}, \frac{q_1-q_0}{h}\right), \\ \frac{p_1 + p_0}{2} &= \frac{\partial L}{\partial \dot{q}}\left(\frac{q_1+q_0}{2}, \frac{q_1-q_0}{h}\right). \end{aligned} \tag{3.22}$$

The second of these equations is simply the statement that

$$\left(\frac{q_1+q_0}{2}, \frac{p_1+p_0}{2}\right) = \mathbb{F}L\left(\frac{q_1+q_0}{2}, \frac{q_1-q_0}{h}\right),$$

and so using (2.17a) shows that (3.22) is equivalent to (3.21b), while (2.17c) gives (3.21a).

For regular Lagrangian systems, the midpoint discrete Lagrangian $L_d^{1/2}$ thus has discrete Hamiltonian map which is the midpoint rule on T^*Q for the corresponding Hamiltonian system.

3.6.2 Störmer-Verlet

The Verlet method [Verlet, 1967] (also known as Störmer's rule) was originally formulated for molecular dynamics problems and remains popular in that field. The derivation of Verlet as a variational integrator is in Wendlandt and Marsden [1997a] and is implicitly in Gillilan and Wilson [1992] as well.

Verlet is usually written for systems of the form $L(q,\dot{q}) = \frac{1}{2}\dot{q}^T M\dot{q} - V(q)$, and was originally formulated as a map $Q \times Q \to Q \times Q$ given by $(q_k, q_{k+1}) \mapsto (q_{k+1}, q_{k+2})$ with

$$q_{k+1} = 2q_k - q_{k-1} + h^2 a_k,$$

where we use the notation $a_k = M^{-1}(-\nabla V(q_k))$. As can be readily seen, this is just the discrete Lagrangian map $F_{L_d} : Q \times Q \to Q \times Q$ for either of

$$\begin{aligned} L_d^0(q_0, q_1, h) &= hL\left(q_0, \frac{q_1-q_0}{h}\right), \\ L_d^1(q_0, q_1, h) &= hL\left(q_1, \frac{q_1-q_0}{h}\right), \end{aligned}$$

or indeed any affine combination of these two. In particular, consider the symmetric version

$$L_d(q_0, q_1, h) = \frac{1}{2}hL\left(q_0, \frac{q_1-q_0}{h}\right) + \frac{1}{2}hL\left(q_1, \frac{q_1-q_0}{h}\right),$$

which gives Verlet as the corresponding F_{L_d}. Pushing this forward to T^*Q with $\mathbb{F}^{\pm}L_d$ now gives

$\tilde{F}_{L_d} : T^*Q \to T^*Q$ defined by (2.23). Evaluating these yields

$$p_k = M\left(\frac{q_{k+1} - q_k}{h}\right) + \frac{1}{2}h\nabla V(q_k),$$
$$p_{k+1} = M\left(\frac{q_{k+1} - q_k}{h}\right) - \frac{1}{2}h\nabla V(q_{k+1}).$$

Now we subtract the first equation from the second and solve the first equation for q_{k+1} to obtain

$$q_{k+1} = q_k + hM^{-1}p_k + \frac{1}{2}h^2M^{-1}(-\nabla V(q_k)),$$
$$p_{k+1} = p_k + h\left(\frac{-\nabla V(q_k) - \nabla V(q_{k+1})}{2}\right),$$

which is the so-called ***velocity Verlet*** method [Swope, Andersen, Berens, and Wilson, 1982; Allen and Tildesley, 1987] written on T^*Q. Using the Legendre transform $\mathbb{F}L(q, \dot{q}) = (q, M\dot{q})$ this can also be mapped to TQ.

We thus see that velocity Verlet will preserve the canonical two-form Ω on T^*Q, and as L_d is invariant under linear symmetries of the potential, Verlet will also preserve quadratic momentum maps such as linear and angular momentum.

3.6.3 Newmark methods

The Newmark family of integrators, originally given in Newmark [1959], are widely used in structural dynamics codes. They are usually written (see, for example, Hughes [1987]) for the system $L = \frac{1}{2}\dot{q}^T M\dot{q} - V(q)$ as maps $TQ \to TQ$ given by $(q_k, \dot{q}_k) \mapsto (q_{k+1}, \dot{q}_{k+1})$ satisfying the implicit relations

$$q_{k+1} = q_k + h\dot{q}_k + \frac{h^2}{2}\left[(1-2\beta)a(q_k) + 2\beta a(q_{k+1})\right], \tag{3.23a}$$
$$\dot{q}_{k+1} = \dot{q}_k + h\left[(1-\gamma)a(q_k) + \gamma a(q_{k+1})\right], \tag{3.23b}$$
$$a(q) = M^{-1}(-\nabla V(q)), \tag{3.23c}$$

where the parameters $\gamma \in [0, 1]$ and $\beta \in [0, \frac{1}{2}]$ specify the method. It is simple to check that the method is second-order if $\gamma = 1/2$ and first-order otherwise, and that it is generally explicit only for $\beta = 0$.

The $\beta = 0$, $\gamma = 1/2$ case is well known to be symplectic (see, for example, Simo et al. [1992]) with respect to the canonical symplectic form Ω_L on TQ. This can be easily seen from the fact that this method is simply the pullback by $\mathbb{F}L$ of the discrete Hamiltonian map for $L_d^{\mathrm{sym},\alpha}$ with $\alpha = 0$ or $\alpha = 1$. Note that this method is the same as velocity Verlet.

It is also well known (for example, Simo et al. [1992]) that the Newmark algorithm with $\beta \neq 0$ does not preserve the *canonical* symplectic form. Nonetheless, based on a remark by Suris, it can

be shown [Kane et al., 2000] that the Newmark method with $\gamma = 1/2$ and any β can be generated from a discrete Lagrangian. To see this, we introduce the map $\varphi^\beta : Q \times Q \to TQ$ defined by

$$\varphi^\beta(q_k, q_{k+1}) = \left(q_k, \left[\frac{q_{k+1} - q_k}{h}\right] - \frac{h}{2}\left[(1-2\beta)a(q_k) + 2\beta a(q_{k+1})\right]\right).$$

Pulling the Newmark map back by φ^β to a map $Q \times Q \to Q \times Q$ gives the map $(q_k, q_{k+1}) \mapsto (q_{k+1}, q_{k+2})$ where

$$\frac{q_{k+2} - 2q_{k+1} + q_k}{h^2} = \beta a(q_{k+2}) + (1-2\beta)a(q_{k+1}) + \beta a(q_k). \tag{3.24}$$

A straightforward calculation now shows that this is in fact the discrete Lagrange map $F_{L_d^\beta}$ for the discrete Lagrangian

$$L_d^\beta(q_0, q_1, h) = h\frac{1}{2}\left(\frac{\eta^\beta(q_1) - \eta^\beta(q_0)}{h}\right)^T M \left(\frac{\eta^\beta(q_1) - \eta^\beta(q_0)}{h}\right) - h\tilde{V}(\eta^\beta(q_0)),$$

where we have introduced the map $\eta^\beta : Q \to Q$ defined by

$$\eta^\beta(q) = q - \beta h^2 M^{-1} \nabla V(q)$$

and the modified potential function $\tilde{V} : Q \to \mathbb{R}$ satisfying $\nabla\tilde{V} \circ \eta^\beta = \nabla V$, which will exist for small h.

This result shows that the Newmark method for $\gamma = 1/2$ is the pullback of the discrete Hamiltonian map $\tilde{F}_{L_d^\beta}$ by the map $\mathbb{F}^+L_d^\beta \circ (\varphi^\beta)^{-1}$. As the discrete Hamiltonian map preserves the canonical symplectic form on T^*Q, this means that Newmark preserves the two-form $[\mathbb{F}^+L_d^\beta \circ (\varphi^\beta)^{-1}]^*\Omega$ on TQ. Note that this is not the canonical two-form Ω_L on TQ, but this is enough to explain the otherwise inexplicably good longtime behaviour of $\gamma = 1/2$ Newmark for nonlinear problems.

An alternative and independent method of analyzing the symplectic members of Newmark has been given by Skeel et al. [1997], including an interesting nonlinear analysis in Skeel and Srinivas [2000]. This is based on the observation that if we define the map $\bar{\eta}^\beta : TQ \to TQ$ by

$$\bar{\eta}^\beta(q, v) = \big(\eta^\beta(q), v\big),$$

then the pushforward of the Newmark method by $\bar{\eta}^\beta$ is given by $(x_k, v_k) \mapsto (x_{k+1}, v_{k+1})$, where

$$x_{k+1} = x_k + hv_k + \frac{1}{2}h^2 a_k, \tag{3.25a}$$

$$v_{k+1} = v_k + \frac{1}{2}h(a_k + a_{k+1}), \tag{3.25b}$$

$$a_k = a(x_k + \beta h^2 a_k). \tag{3.25c}$$

This map can be shown to be symplectic with respect to the canonical two-form Ω_L on TQ, and so Newmark will preserve the two-form $(\bar{\eta}^\beta)^*\Omega_L$ on TQ.

To summarize, we have the following commutative diagram, where the map $\tilde{F}_{L_d^\beta}$ preserves the canonical two-form Ω on T^*Q, the map (3.25) preserves the Lagrange two-form Ω_L on TQ, and we have set $\gamma = 1/2$ in the Newmark equation (3.23).

$$\begin{array}{ccccccc}
T^*Q & \xleftarrow{\mathbb{F}^+L_d^\beta} & Q\times Q & \xrightarrow{\varphi^\beta} & TQ & \xrightarrow{\bar{\eta}^\beta} & TQ \\
\downarrow{\scriptstyle \tilde{F}_{L_d^\beta}} & & \downarrow{\scriptstyle F_{L_d^\beta}}\ (3.24) & & \downarrow{\scriptstyle (3.23)} & & \downarrow{\scriptstyle (3.25)} \\
T^*Q & \xleftarrow[\mathbb{F}^+L_d^\beta]{} & Q\times Q & \xrightarrow[\varphi^\beta]{} & TQ & \xrightarrow[\bar{\eta}^\beta]{} & TQ
\end{array}$$

3.6.4 Explicit symplectic partitioned Runge-Kutta methods

Symplectic integrators which are explicit partitioned Runge-Kutta methods were first used by Ruth [1983] and Forest and Ruth [1990], who constructed them as a composition of steps, each one generated by a generating function of the third kind. Using the same idea shows that these methods are also variational, at least for Hamiltonians with kinetic energy of the form $T(p) = 1/2p^T M^{-1}p$ for some constant mass matrix M.

It can be shown [Hairer et al., 1993] that explicit symplectic partitioned Runge-Kutta methods can always be written as the composition of a number of steps of the method $F^{a,b} : T^*Q \times \mathbb{R} \to T^*Q$ given by

$$\begin{aligned} q_1 &= q_0 + ahM^{-1}p_0, \\ p_1 &= p_0 - bh\nabla V(q_1), \end{aligned}$$

and of its adjoint method $(F^{a,b})^*$, with each step having different parameters (a, b). Furthermore, it is simple to check that these can be chosen so that all steps have nonzero a.

We now see, however, that the method $F^{a,b}$ is the discrete Hamiltonian map for the discrete Lagrangian $L_d^{a,b}$ given by

$$L_d^{a,b}(q_0, q_1, h) = h\left[b\frac{1}{2}\left(\frac{q_1 - q_0}{h}\right)^T M\left(\frac{q_1 - q_0}{h}\right) - \frac{1}{a}V(q_1)\right],$$

and from Theorem 3.4 it is clear that $(F^{a,b})^*$ is the discrete Hamiltonian map of the adjoint discrete Lagrangian $(L_d^{a,b})^*$.

We can thus form a composition discrete Lagrangian as in Theorem 3.5 whose discrete Hamiltonian map is the composition of the $F^{a,b}$ and $(F^{a,b})^*$, and is therefore the explicit symplectic partitioned Runge-Kutta method.

3.6.5 Symplectic partitioned Runge-Kutta methods

Partitioned Runge-Kutta methods are a class of integrators about which much is known and which generalize standard Runge-Kutta methods. The symplectic members of Runge-Kutta were first identified by Lasagni [1988], Sanz-Serna [1988] and Suris [1989]. Symplectic partitioned Runge-Kutta methods appeared in Sanz-Serna [1992a] and Suris [1990]. Good general references are Hairer et al. [1993] and Hairer and Wanner [1996]. See also Geng [1995, 2000], Sofroniou and Oevel [1997a] and Sofroniou and Oevel [1997b] for order conditions and derivations. An explicit construction has been given by Suris [1990] for the discrete Lagrangian which generates any symplectic partitioned Runge-Kutta method. We summarize this derivation below.

Recall that a partitioned Runge-Kutta method for the regular Lagrangian system L is a map $T^*Q \times \mathbb{R} \to T^*Q$ specified by the coefficients b_i, a_{ij}, $\tilde{b}_i$, $\tilde{a}_{ij}$ for $i, j = 1, \ldots, s$, and defined by $(q_0, p_0) \mapsto (q_1, p_1)$ for

$$q_1 = q_0 + h \sum_{j=1}^{s} b_j \dot{Q}_j, \qquad p_1 = p_0 + h \sum_{j=1}^{s} \tilde{b}_j \dot{P}_j, \tag{3.26a}$$

$$Q_i = q_0 + h \sum_{j=1}^{s} a_{ij} \dot{Q}_j, \qquad P_i = p_0 + h \sum_{j=1}^{s} \tilde{a}_{ij} \dot{P}_j, \qquad i = 1, \ldots, s, \tag{3.26b}$$

$$P_i = \frac{\partial L}{\partial \dot{q}}(Q_i, \dot{Q}_i), \qquad \dot{P}_i = \frac{\partial L}{\partial q}(Q_i, \dot{Q}_i), \qquad i = 1, \ldots, s, \tag{3.26c}$$

where the points (Q_i, P_i) are known as the internal stages. In the special case that $a_{ij} = \tilde{a}_{ij}$ and $b_i = \tilde{b}_i$, a partitioned Runge-Kutta method is said to be simply a Runge-Kutta method.

It is well known that the method is symplectic (that is, it preserves the canonical symplectic form Ω on T^*Q) if the coefficients satisfy

$$b_i \tilde{a}_{ij} + \tilde{b}_j a_{ji} = b_i \tilde{b}_j, \qquad i, j = 1, \ldots, s, \tag{3.27a}$$

$$b_i = \tilde{b}_i, \qquad i = 1, \ldots, s. \tag{3.27b}$$

We now assume that we have coefficients satisfying (3.27) and write a discrete Lagrangian that generates the corresponding symplectic partitioned Runge-Kutta method. Given points $(q_0, q_1) \in Q \times Q$, we can regard (3.26) as implicitly defining p_0, p_1, Q_i, P_i, $\dot{Q}_i$ and $\dot{P}_i$ for $i = 1, \ldots, s$. Taking these to be so defined as functions of (q_0, q_1), we construct a discrete Lagrangian

$$L_d(q_0, q_1, h) = h \sum_{i=1}^{s} b_i L(Q_i, \dot{Q}_i). \tag{3.28}$$

It can now be shown [Suris, 1990] that the corresponding discrete Hamiltonian map is exactly the map $(q_0, p_0) \mapsto (q_1, p_1)$, which is the symplectic partitioned Runge-Kutta method. Nonsymplectic

partitioned Runge-Kutta methods will clearly not have a corresponding discrete Lagrangian formulation.

Theorem 3.6. *The discrete Hamiltonian map generated by the discrete Lagrangian (3.28) is a symplectic partitioned Runge-Kutta method.*

Proof. To check that the discrete Hamiltonian map defined by L_d is indeed the partitioned Runge-Kutta method specified by (3.26), we need only check that equations (2.23) are satisfied. We compute

$$\begin{aligned}\frac{\partial L_d}{\partial q_0}(q_0,q_1) &= (\Delta t)\sum_{i=1}^{s} b_i\left[\frac{\partial L}{\partial q}\cdot\frac{\partial Q_i}{\partial q_0}+\frac{\partial L}{\partial \dot{q}}\cdot\frac{\partial \dot{Q}_i}{\partial q_0}\right]\\ &= (\Delta t)\sum_{i=1}^{s} b_i\left[\dot{P}_i\cdot\frac{\partial Q_i}{\partial q_0}+P_i\cdot\frac{\partial \dot{Q}_i}{\partial q_0}\right],\end{aligned}$$

using the definitions for P_i and $\dot{P}_i$ in (3.26). Differentiating the definition for Q_i in (3.26b) and substituting in this and the definition of P_i in (3.26b) now gives

$$\begin{aligned}\frac{\partial L_d}{\partial q_0}(q_0,q_1) &= (\Delta t)\sum_{i=1}^{s} b_i\left[\dot{P}_i\cdot\left(I+(\Delta t)\sum_{j=1}^{s}a_{ij}\frac{\partial \dot{Q}_j}{\partial q_0}\right)\right.\\ &\qquad\left.+\left(p_0+(\Delta t)\sum_{j=1}^{s}\tilde{a}_{ij}\dot{P}_j\right)\cdot\frac{\partial \dot{Q}_i}{\partial q_0}\right]\\ &= (\Delta t)\sum_{i=1}^{s} b_i\left[\dot{P}_i+p_0\cdot\frac{\partial \dot{Q}_i}{\partial q_0}\right]\\ &\qquad+(\Delta t)^2\sum_{i=1}^{s}\sum_{j=1}^{s}(b_i\tilde{a}_{ij}+b_j a_{ji})\dot{P}_j\cdot\frac{\partial \dot{Q}_i}{\partial q_0},\end{aligned}$$

and we can use the symplecticity identities (3.27) to obtain

$$\begin{aligned}\frac{\partial L_d}{\partial q_0}(q_0,q_1) &= p_0\cdot\left[(\Delta t)\sum_{i=1}^{s} b_i\frac{\partial \dot{Q}_i}{\partial q_0}\right]+(\Delta t)\sum_{i=1}^{s} b_i\dot{P}_i\\ &\qquad+(\Delta t)\sum_{j=1}^{s} b_j\dot{P}_j\cdot\left[(\Delta t)\sum_{i=1}^{s} b_i\frac{\partial \dot{Q}_i}{\partial q_0}\right]\\ &= -p_0,\end{aligned}$$

where we have differentiated the expression for q_1 in (3.26a) to obtain the identity

$$(\Delta t)\sum_{i=1}^{s} b_i\frac{\partial \dot{Q}_i}{\partial q_0} = -I.$$

This thus establishes that the first equation of (2.23) is satisfied.

Differentiating L_d with respect to q_1 and following a similar argument to that above gives the

second part of (2.23), and shows that the discrete Hamiltonian map $\tilde{F}L_d$ generated by the discrete Lagrangian (3.28) is indeed the symplectic partitioned Runge-Kutta method. □

This construction thus provides a proof of the well-known fact that the restrictions (3.27) on the coefficients mean that the partitioned Runge-Kutta method is symplectic, as discrete Hamiltonian maps always preserve the canonical symplectic form. In addition, the linear nature of the definition of the discrete Lagrangian (3.28) means that it will inherit linear symmetries of the Lagrangian L, which thus proves the standard result that partitioned Runge-Kutta methods preserve quadratic momentum maps.

Another way to regard the above derivation is to say that we have written down a generating function of the first kind for the symplectic partitioned Runge-Kutta map.

3.6.6 Galerkin methods

To obtain accurate variational integrators, the results in §3.3 show that the discrete Lagrangian should approximate the action over short trajectory segments. One way to do this practically is to use polynomial approximations to the trajectories and numerical quadrature to approximate the integral. This can be shown to be equivalent both to a class of continuous Galerkin methods and to a subset of symplectic partitioned Runge-Kutta methods.

This approach is related to the Continuous Galerkin and Discontinuous Galerkin methods, as in Estep and French [1994], Hulme [1972b,a] and Thomée [1997]. These methods differ in the precise choice of function space (continuous or discontinuous) and whether the position and velocity components are projected separately or the velocity projection is given by the lift of a position projection.

We know that a discrete Lagrangian should be an approximation

$$L_d(q_0, q_1, h) \approx \operatorname*{ext}_{q \in \mathcal{C}([0,h],Q)} \mathfrak{G}(q),$$

where $\mathcal{C}([0,h],Q)$ is the space of trajectories $q : [0,h] \to Q$ with $q(0) = q_0$ and $q(h) = q_1$, and $\mathfrak{G} : \mathcal{C}(0,h) \to \mathbb{R}$ is the action (2.1).

To approximate this quantity, we choose the particular finite-dimensional approximation $\mathcal{C}^s([0,h],Q) \subset \mathcal{C}([0,h],Q)$ of the trajectory space given by

$$\mathcal{C}^s([0,h],Q) = \{q \in \mathcal{C}([0,h],Q) \mid q \text{ is a polynomial of degree } s\},$$

and we approximate the action integral with numerical quadrature to give an approximate action

$\mathfrak{G}^s : \mathcal{C}([0,h],Q) \to \mathbb{R}$ by

$$\mathfrak{G}^s(q) = h \sum_{i=1}^{s} b_i L\big(q(c_i h), \dot{q}(c_i h)\big), \tag{3.29}$$

where $c_i \in [0,1]$, $i = 1, \ldots, s$ are a set of quadrature points and b_i are the associated maximal order weights. We now set the Galerkin discrete Lagrangian to be

$$L_d(q_0, q_1, h) = \operatorname*{ext}_{q \in \mathcal{C}^s([0,h],Q)} \mathfrak{G}^s(q), \tag{3.30}$$

which can be practically evaluated. This procedure, of course, is simply performing Galerkin projection of the weak form of the ODE onto the space of piecewise polynomial trajectories. Furthermore, as we will show below, the resulting integrator is a symplectic partitioned Runge-Kutta method.

To make the above equations explicit, choose control times $0 = d_0 < d_1 < d_2 < \cdots < d_{s-1} < d_s = 1$ and control points $q_0^0 = q_0, q_0^1, q_0^2, \ldots, q_0^{s-1}, q_0^s = q_1$. These uniquely define the degree s polynomial $q_d(t; q_0^\nu, h)$ which passes through each q_0^ν at time $d_\nu h$, that is, $q_d(d_\nu h) = q_0^\nu$ for $\nu = 0, \ldots, s$. Letting $\tilde{l}_{\nu,s}(t)$ denote the Lagrange polynomials associated with the d_ν, we can express $q_d(t; q_0^\nu, h)$ as

$$q_d(\tau h; q_0^\nu, h) = \sum_{\nu=0}^{s} q_0^\nu \tilde{l}_{\nu,s}(\tau). \tag{3.31}$$

For $q_d(t; q_0^\nu, h)$ to be a critical point of the discrete action (3.29) we must have stationarity with respect to variations in q_0^ν for $\nu = 1, \ldots, s-1$. Differentiating (3.29) and (3.31) implies that we have

$$0 = h \sum_{i=1}^{s} b_i \left[\frac{\partial L}{\partial q}(c_i h) \tilde{l}_{\nu,s}(c_i) + \frac{1}{h} \frac{\partial L}{\partial \dot{q}}(c_i h) \dot{\tilde{l}}_{\nu,s}(c_i) \right] \tag{3.32}$$

for each $\nu = 1, \ldots, s-1$, where we denote $\frac{\partial L}{\partial q}(c_i h) = \frac{\partial L}{\partial q}(q_d(c_i h), \dot{q}_d(c_i h))$ and similarly for the other expressions.

The integration scheme $(q_0, p_0) \mapsto (q_1, p_1)$ generated by the Galerkin discrete Lagrangian (3.30) is now given implicitly by the relations

$$-p_0 = \frac{\partial L_d}{\partial q_0}(q_0, q_1, h), \qquad p_1 = \frac{\partial L_d}{\partial q_1}(q_0, q_1, h).$$

Evaluating these expressions and restating (3.32) gives the set of equations

$$\begin{aligned}
E(0) : -p_0 &= h \sum_{i=1}^{s} b_i \left[\frac{\partial L}{\partial q}(c_i h) \tilde{l}_{0,s}(c_i) + \frac{1}{h} \frac{\partial L}{\partial \dot{q}}(c_i h) \dot{\tilde{l}}_{0,s}(c_i) \right], \\
E(\nu) : \quad 0 &= h \sum_{i=1}^{s} b_i \left[\frac{\partial L}{\partial q}(c_i h) \tilde{l}_{\nu,s}(c_i) + \frac{1}{h} \frac{\partial L}{\partial \dot{q}}(c_i h) \dot{\tilde{l}}_{\nu,s}(c_i) \right], \ \nu = 1, \ldots, s-1, \\
E(s) : \quad p_1 &= h \sum_{i=1}^{s} b_i \left[\frac{\partial L}{\partial q}(c_i h) \tilde{l}_{s,s}(c_i) + \frac{1}{h} \frac{\partial L}{\partial \dot{q}}(c_i h) \dot{\tilde{l}}_{s,s}(c_i) \right],
\end{aligned}$$

which define the discrete Hamiltonian map $(q_0, p_0) \mapsto (q_1, p_1)$.

The above Galerkin discrete Lagrangian can also be interpreted as a function of several points, in a similar way to the composition discrete Lagrangians discussed in §3.5. Essentially, we choose a set of interior points which act as a parametrization of the space of degree s polynomials mapping $[0, h]$ to Q.

More precisely, we form the multipoint discrete Lagrangian

$$L_d(q_0^0, q_0^1, \ldots, q_0^s, h) = \mathfrak{G}^s\big(q_d(t; q_0^\nu, h)\big),$$

where we recall that $q_d(t; q_0^\nu, h)$ is the unique polynomial of degree s passing through q_0^ν at time $d_\nu h$ and $\mathfrak{G}^s$ is defined by (3.29). This multipoint discrete Lagrangian is the analogue of the discrete Lagrangian (3.17). The appropriate discrete action is then

$$\mathfrak{G}_d\big(\{(q_k = q_k^0, \ldots, q_k^s = q_{k+1})\}_{k=1}^N\big) = \sum_{k=0}^{N} L_d(q_k^0, q_k^1, \ldots, q_k^s, h),$$

and the corresponding discrete Euler-Lagrange equations are given by (3.19). Clearly, the discrete Lagrangian defined by extremizing the above multipoint L_d with respect to the interior points q_0^ν for $\nu = 1, \ldots, s$ is just the original Galerkin discrete Lagrangian (3.30), and the extended discrete Euler-Lagrange equations are thus equivalent to $E(\nu)$ above for $\nu = 0, \ldots, s$. This follows in the same way as the proof of Theorem 3.5.

We will now see that these Galerkin variational integrators can be realized as particular examples of Runge-Kutta or partitioned Runge-Kutta schemes.

Theorem 3.7. *Take a set of quadrature points c_i with corresponding maximal order weights b_i and let L_d be the corresponding Galerkin discrete Lagrangian (3.30). Then the integrator generated by*

this discrete Lagrangian is the partitioned Runge-Kutta scheme defined by the coefficients

$$\begin{aligned} b_i &= \tilde{b}_i = \int_0^1 l_{i,s}(\rho)\mathbf{d}\rho, \\ a_{ij} &= \int_0^{c_i} l_{j,s}(\rho)\mathbf{d}\rho, \\ \tilde{a}_{ij} &= \tilde{b}_j \left(1 - \frac{a_{ji}}{b_i}\right), \end{aligned} \tag{3.33}$$

where the $l_{i,s}(\rho)$ are the Lagrange polynomials associated with the c_i.

Proof. Given (q_0, p_0), (q_1, p_1) and q_0^ν satisfying $E(\nu)$, $\nu = 0, \ldots, s$, we will show that they also satisfy the partitioned Runge-Kutta equations (3.26) written for a Lagrangian system with coefficients given by (3.33). We restate the defining equations here for reference:

$$q_1 = q_0 + h\sum_{j=1}^{s} b_j \dot{Q}_j, \qquad p_1 = p_0 + h\sum_{j=1}^{s} \tilde{b}_j \dot{P}_j, \tag{3.34a}$$

$$Q_i = q_0 + h\sum_{j=1}^{s} a_{ij} \dot{Q}_j, \qquad P_i = p_0 + h\sum_{j=1}^{s} \tilde{a}_{ij} \dot{P}_j, \qquad i = 1, \ldots, s, \tag{3.34b}$$

$$P_i = \frac{\partial L}{\partial \dot{q}}(Q_i, \dot{Q}_i), \qquad \dot{P}_i = \frac{\partial L}{\partial q}(Q_i, \dot{Q}_i), \qquad i = 1, \ldots, s. \tag{3.34c}$$

We will show that these equations are satisfied by the discrete Hamiltonian map.

Set $\dot{Q}_i = \dot{q}_d(c_i h; q_0^\nu, h)$ so that $\dot{q}_d(\tau h; q_0^\nu, h) = \sum_{j=1}^s \dot{Q}_j l_{j,s}(\tau)$. Integrating this expression and using the fact that $q_d(0; q_0^\nu, h) = q_0$ gives

$$q_d(\tau h; q_0^\nu, h) = q_0 + h\sum_{j=1}^{s} \dot{Q}_j \int_0^\tau l_j(\rho)\mathbf{d}\rho.$$

Setting $Q_i = q_d(c_i h; q_0^\nu, h)$ and using $q_1 = q_d(h; q_0^\nu, h)$ now gives the first parts of (3.34a) and (3.34b) for Q_i and q_1. Now define P_i and $\dot{P}_i$ according to (3.34c).

Until this point we have not made use of the relations $E(\nu)$. We will now begin to do so by forming the sum of the $E(\nu)$, $\nu = 0, \ldots, s$. This gives

$$p_1 - p_0 = h\sum_{i=1}^{s} b_i \left[\frac{\partial L}{\partial q}(c_i h)\sum_{\nu=0}^{s} \tilde{l}_{\nu,s}(c_i) + \frac{1}{h}\frac{\partial L}{\partial \dot{q}}(c_i h)\sum_{\nu=0}^{s} \dot{\tilde{l}}_{\nu,s}(c_i)\right].$$

However, the Lagrange polynomials $\tilde{l}_{\nu,s}(\tau)$ sum to the identity function, and therefore the sum of their derivatives must be zero. We thus recover the second part of (3.34a) for p_1.

Note that the $\tilde{l}_{\nu,s+1}$ are a set of $s+1$ independent polynomials of degree s and thus are a basis for $\mathcal{P}_s$, the space of polynomials of degree s. For each $j = 1, \ldots, s$ the polynomial $l_{j,s}$ is of degree

$s-1$ and so has an integral of degree s. This implies that there exist coefficients m_ν^j such that

$$\sum_{\nu=0}^{s} m_\nu^j \tilde{l}_{\nu,s+1}(\tau) = \int_0^\tau l_{j,s}(\rho)\mathbf{d}\rho - b_j.$$

Differentiating this expression with respect to τ and evaluating it at $\tau = 0$ and $\tau = 1$ gives the following three identities:

$$\begin{aligned}
\sum_{\nu=0}^{s} m_\nu^j \dot{\tilde{l}}_{\nu,s+1}(\tau) &= l_{j,s}(\tau), \\
m_s^j &= \sum_{\nu=0}^{s} m_\nu^j \tilde{l}_{\nu,s+1}(1) = \int_0^1 l_{j,s}(\rho)\mathbf{d}\rho - b_j = 0, \\
m_0^j &= \sum_{\nu=0}^{s} m_\nu^j \tilde{l}_{\nu,s+1}(0) = -b_j.
\end{aligned}$$

If we now form the sum $\sum_{\nu=0}^{s} m_\nu^j E(\nu)$ and make use of the above identities, we obtain

$$\begin{aligned}
b_j p_0 &= h \sum_{i=1}^{s} b_i \left[\dot{P}_i \left(\int_0^{c_i} l_{j,s}(\rho)\mathbf{d}\rho - b_j \right) + \frac{1}{h} P_i l_{j,s}(c_i) \right] \\
&= h \sum_{i=1}^{s} \dot{P}_i \left[b_i (a_{ij} - b_j) \right] + b_j P_j,
\end{aligned}$$

which can be rearranged to give the second part of (3.34b) for P_i. □

If the $\hat{a}_{ij}$ in Theorem 3.7 are equal to the a_{ij}, then the method is clearly the special case of a Runge-Kutta method, rather than the general partitioned Runge-Kutta case. Note that the definition of the $\hat{a}_{ij}$ in (3.33) is simply a rearrangement of the requirement (3.27a), and so the partitioned Runge-Kutta methods equivalent to the Galerkin variational integrators are naturally symplectic, as is clear from the symplectic nature of variational integrators in general. In addition, the additive structure of the Galerkin discrete Lagrangian means that L_d will inherit linear symmetries of L, so Noether's theorem recovers the well-known fact that the partitioned Runge-Kutta methods will preserve quadratic momentum maps.

A particularly elegant symplectic Runge-Kutta method is the ***collocation Gauss-Legendre rule***. In the present derivation this results simply from taking the quadrature points c_i to be those given by the Gauss-Legendre quadrature, which is the highest-order quadrature for a given number of points. The c_i produced in this manner are all strictly between 0 and 1.

If the system being integrated is stiff, then better numerical performance results from having $c_s = 1$, making the integrator stiffly accurate [Hairer and Wanner, 1996]. If we also wish to preserve the symmetry of the discrete Lagrangian, then it is natural to seek the c_i giving the highest order quadrature rule while enforcing $c_0 = 0$ and $c_s = 1$. This is the so-called Lobatto quadrature, and

the Galerkin variational integrator generated in this way is the ***standard Lobatto IIIA-IIIB partitioned Runge-Kutta method***.

Chapter 4

Forcing and constraints

4.1 Background: Forced systems

Lagrangian and Hamiltonian systems with external forcing arise in many different contexts. Particular examples include control forces from actuators, dissipation and friction, and loading on mechanical systems. As we will see below, when integrating such systems it is important to take account of the geometric structure to avoid spurious numerical artifacts. One way to do this is by extending the discrete variational framework to include forcing, which we will now do.

4.1.1 Forced Lagrangian systems

A ***Lagrangian force*** is a fibre-preserving map $f_L : TQ \to T^*Q$ over the identity, which we write in coordinates as

$$f_L : (q, \dot{q}) \mapsto (q, f_L(q, \dot{q})).$$

Given such a force, it is standard to modify Hamilton's principle, seeking stationary points of the action, to the ***Lagrange-d'Alembert principle***, which seeks curves $q \in \mathcal{C}(Q)$ satisfying

$$\delta \int_0^T L(q(t), \dot{q}(t))dt + \int_0^T f_L(q(t), \dot{q}(t)) \cdot \delta q(t) dt = 0, \tag{4.1}$$

where the δ represents variations vanishing at the endpoints. Using integration by parts shows that this is equivalent to the ***forced Euler-Lagrange equations***, which have coordinate expression

$$\frac{\partial L}{\partial q}(q, \dot{q}) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}(q, \dot{q})\right) + f_L(q, \dot{q}) = 0. \tag{4.2}$$

Note that these are the same as the standard Euler-Lagrange equations (2.5) with the forcing term added.

4.1.2 Forced Hamiltonian systems

A ***Hamiltonian force*** is a fibre-preserving map $f_H : T^*Q \to T^*Q$ over the identity. Given such a force, we define the corresponding horizontal one-form f'_H on T^*Q by

$$f'_H(p_q) \cdot u_{p_q} = \left\langle f_H(p_q), T\pi_Q \cdot u_{p_q} \right\rangle,$$

where $\pi_Q : T^*Q \to Q$ is the projection. This expression is reminiscent of the definition (2.13) of the canonical one-form Θ on T^*Q, and in coordinates it is $f'_H(q,p) \cdot (\delta q, \delta p) = f_H(q,p) \cdot \delta q$, so the one-form is clearly horizontal.

The ***forced Hamiltonian vector field*** X_H is now defined to satisfy

$$\mathbf{i}_{X_H}\Omega = \mathbf{d}H - f'_H$$

and in coordinates this gives the well-known ***forced Hamilton's equations***

$$X_q(q,p) = \frac{\partial H}{\partial q}(q,p), \tag{4.3a}$$

$$X_p(q,p) = -\frac{\partial H}{\partial p}(q,p) + f_H(q,p), \tag{4.3b}$$

which are the same as the standard Hamilton's equations (2.15) with the forcing term added to the momentum equation.

4.1.3 Legendre transform with forces

Given a Lagrangian L, we can take the standard Legendre transform $\mathbb{F}L : T^*Q \to TQ$ and relate Hamiltonian and Lagrangian forces by

$$f_L = f_H \circ \mathbb{F}L.$$

If we also have a Hamiltonian H related to L by the Legendre transform according to (2.16), then it can be shown that the forced Euler-Lagrange equations and the forced Hamilton's equations are equivalent. That is, if X_L and X_H are the forced Lagrangian and Hamiltonian vector fields, respectively, then $(\mathbb{F}L)^*(X_H) = X_L$.

4.1.4 Noether's theorem with forcing

We now consider the effect of forcing on the evolution of momentum maps that arise from symmetries of the Lagrangian $L : TQ \to \mathbb{R}$. Let $\Phi : G \times Q \to Q$ be a symmetry of L and let the Lagangian momentum map $J_L : TQ \to \mathfrak{g}^*$ be as defined in Section 2.1.4.

Evaluating the left-hand side of (4.1) for a variation of the form $\delta q(t) = \xi_Q(q(t))$ gives

$$\int_0^T \mathbf{d}L \cdot \xi_{TQ} dt + \int_0^T f_L \cdot \xi_Q dt = \int_0^T f_L \cdot \xi_Q dt,$$

as L is assumed to be invariant. Using integration by parts as in the derivation of the forced Euler-Lagrange equations, we see that the above expression is equal to

$$\int_0^T \left[\frac{\partial L}{\partial q}(q, \dot{q}) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}(q, \dot{q}) \right) + f_L(q, \dot{q}) \right] + \Theta_L \cdot \xi_{TQ}|_0^T$$
$$= (J_L \circ F_L^T)(q(0), \dot{q}(0)) \cdot \xi - J_L(q(0), \dot{q}(0)) \cdot \xi,$$

and so equating these two statements of (4.1) gives

$$\left[(J_L \circ F_L^T)(q(0), \dot{q}(0)) - J_L(q(0), \dot{q}(0)) \right] \cdot \xi = \int_0^T f_L(q(t), \dot{q}(t)) \cdot \xi_Q(q(t)) dt. \tag{4.4}$$

This equation describes the evolution of the momentum map from time 0 to time T, and shows that forcing will generally alter the momentum map. In the special case that the forcing is orthogonal to the group action, the above derivation shows that Noether's theorem will still hold.

Theorem 4.1 (Forced Noether's theorem). *Consider a Lagrangian system $L : TQ \to \mathbb{R}$ with forcing $f_L : TQ \to T^*Q$ and a symmetry action $\Phi : G \times Q \to Q$ such that $\langle f_L(q, \dot{q}), \xi_Q(q) \rangle = 0$ for all $(q, \dot{q}) \in TQ$ and all $\xi \in \mathfrak{g}$. Then the Lagrangian momentum map $J_L : TQ \to \mathfrak{g}^*$ will be preserved by the flow, so that $J_L \circ F_L^t = J_L$ for all t.*

A similar result can also be derived for Hamiltonian systems, either by taking the Legendre transform of a regular forced Lagrangian system, or by working directly on the Hamiltonian side as in Section 2.3. For more details on the relationship between momentum maps and forcing see Bloch, Krishnaprasad, Marsden, and Ratiu [1996].

Note that, for nonzero forcing, the Lagrangian and Hamiltonian flows do not preserve the symplectic two-form. This can be seen by calculating $\mathbf{d}\hat{\mathfrak{G}}$ as was done in Section 2.1.3, and realizing that it contains a term with the integral of the force which does not vanish except when $f_L = 0$.

4.2 Discrete variational mechanics with forces

4.2.1 Discrete Lagrange-d'Alembert principle

As with other discrete structures, we take two ***discrete Lagrangian forces*** $f_d^+, f_d^- : Q \times Q \to T^*Q$, which are fibre-preserving in the sense that $\pi_Q \circ f_d^\pm = \pi_Q^\pm$, and which thus have coordinate expressions

$$f_d^+(q_0, q_1) = (q_1, f_d^+(q_0, q_1)),$$
$$f_d^-(q_0, q_1) = (q_0, f_d^-(q_0, q_1)).$$

We combine the two discrete forces to give a single one-form $f_d : Q \times Q \to T^*(Q \times Q)$ defined by

$$f_d(q_0, q_1) \cdot (\delta q_0, \delta q_1) = f_d^+(q_0, q_1) \cdot \delta q_1 + f_d^-(q_0, q_1) \cdot \delta q_0. \tag{4.5}$$

As with discrete Lagrangians, the discrete forces will also depend on the time step h, which is important when relating the discrete and continuous mechanics. Given such forces, we modify the discrete Hamilton's principle, following Kane et al. [2000], to the ***discrete Lagrange-d'Alembert principle***, which seeks discrete curves $\{q_k\}_{k=0}^N$ that satisfy

$$\delta \sum_{k=0}^{N-1} L_d(q_k, q_{k+1}) + \sum_{k=0}^{N-1} \left[f_d^-(q_k, q_{k+1}) \cdot \delta q_k + f_d^+(q_k, q_{k+1}) \cdot \delta q_{k+1} \right] = 0 \tag{4.6}$$

for all variations $\{\delta q_k\}_{k=0}^N$ vanishing at the endpoints. This is equivalent to the ***forced discrete Euler-Lagrange equations***

$$D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}) + f_d^+(q_{k-1}, q_k) + f_d^-(q_k, q_{k+1}) = 0, \tag{4.7}$$

which are the same as the standard discrete Euler-Lagrange equations (2.11) with the discrete forces added. These implicitly define the ***forced discrete Lagrangian map*** $f_d : Q \times Q \to Q \times Q$.

4.2.2 Discrete Legendre transforms with forces

Although in the continuous case we used the standard Legendre transform for systems with forcing, in the discrete case it is necessary to take the ***forced discrete Legendre transforms*** to be

$$\mathbb{F}^{f+} L_d : (q_0, q_1) \mapsto (q_1, p_1) = (q_1, D_2 L_d(q_0, q_1) + f_d^+(q_0, q_1)), \tag{4.8a}$$
$$\mathbb{F}^{f-} L_d : (q_0, q_1) \mapsto (q_0, p_0) = (q_0, -D_1 L_d(q_0, q_1) - f_d^-(q_0, q_1)). \tag{4.8b}$$

Using these definitions and the forced discrete Euler-Lagrange equations (4.7), we can see that the corresponding ***forced discrete Hamiltonian map*** $\tilde{F}_{L_d} = \mathbb{F}^{f\pm}L_d \circ F_{L_d} \circ (\mathbb{F}^{f\pm}L_d)^{-1}$ is given by the map $\tilde{F}_{L_d} : (q_0, p_0) \mapsto (q_1, p_1)$, where

$$p_0 = -D_1 L_d(q_0, q_1) - f_d^-(q_0, q_1), \tag{4.9a}$$

$$p_1 = D_2 L_d(q_0, q_1) + f_d^+(q_0, q_1), \tag{4.9b}$$

which is the same as the standard discrete Hamiltonian map (2.23) with the discrete forces added.

4.2.3 Discrete Noether's theorem with forcing

Consider a group action $\Phi : G \times Q \to Q$ and assume that the discrete Lagrangian $L_d : Q \times Q \to \mathbb{R}$ is invariant under the lifted product action, as in Section 2.2.3. We can now calculate (4.6) in the direction of a variation $\delta q_k = \xi_Q(q_k)$ to give

$$\begin{aligned}\sum_{k=0}^{N-1} \mathbf{d}L_d(q_k, q_{k+1}) \cdot \xi_{Q\times Q}(q_k, q_{k+1}) + \sum_{k=0}^{N-1} f_d(q_k, q_{k+1}) \cdot \xi_{Q\times Q}(q_k, q_{k+1}) \\ = \sum_{k=0}^{N-1} f_d(q_k, q_{k+1}) \cdot \xi_{Q\times Q}(q_k, q_{k+1}),\end{aligned}$$

or we can use a discrete integration by parts to obtain the alternative expression

$$\begin{aligned}&\sum_{k=1}^{N-1} \big[D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}) + f_d^+(q_{k-1}, q_k) + f_d^-(q_k, q_{k+1})\big] \cdot \xi_Q(q_k) \\ &\quad + \big[D_2 L_d(q_{N-1}, q_N) + f_d^+(q_{N-1}, q_N)\big] \cdot \xi_Q(q_N) \\ &\quad + \big[D_1 L_d(q_0, q_1) + f_d^-(q_0, q_1)\big] \cdot \xi_Q(q_0) \\ &\qquad = \mathbb{F}^{f+}L_d(q_{N-1}, q_N) \cdot \xi_Q(q_N) - \mathbb{F}^{f-}L_d(q_0, q_1) \cdot \xi_Q(q_0).\end{aligned}$$

We now consider how the discrete momentum map should be defined in the presence of forcing, as there is a choice between the expressions (2.7) involving $\Theta_{L_d}^{\pm}$ and the expressions

$$J_{L_d}^{f+}(q_0, q_1) \cdot \xi = \big\langle \mathbb{F}^{f+}L_d(q_0, q_1), \xi_Q(q_1) \big\rangle, \tag{4.10a}$$

$$J_{L_d}^{f-}(q_0, q_1) \cdot \xi = \big\langle \mathbb{F}^{f-}L_d(q_0, q_1), \xi_Q(q_1) \big\rangle, \tag{4.10b}$$

which are based on the discrete Legendre transforms. In the unforced discrete case and in the continuous case both with and without forcing, these expressions are equal to the definition based on Θ_L and so the question does not arise. For a discrete system, however, consideration of the forced exact discrete Lagrangian defined below shows that (4.10) are the correct definitions. Given this,

we can equate the above two forms of (4.6) to obtain

$$\left[J_{L_d}^{f+} \circ F_{L_d}^{N-1} - J_{L_d}^{f-}\right](q_0, q_1) \cdot \xi = \sum_{k=0}^{N-1} f_d(q_k, q_{k+1}) \cdot \xi_{Q\times Q}(q_k, q_{k+1}),$$

which describes the evolution of the discrete momentum map. If the discrete forces are orthogonal to the group action, so that $\langle f_d, \xi_{Q\times Q}\rangle = 0$ for all $\xi \in \mathfrak{g}$, then we have

$$0 = \left\langle \mathbf{d}L_d + f_d, \xi_{Q\times Q}\right\rangle = J_{L_d}^{f+} - J_{L_d}^{f-},$$

and thus the two discrete Lagrangian momentum maps are equal. Denoting this unique map by $J_{L_d}^f : Q \times Q \to \mathfrak{g}^*$, we see that the momentum map evolution equation gives a forced Noether's theorem for discrete mechanics.

Theorem 4.2 (Discrete forced Noether's theorem) *Consider a discrete Lagrangian system $L_d : Q \times Q \to \mathbb{R}$ with discrete forces $f_d^+, f_d^- : Q \times Q \to T^*Q$ and a symmetry action $\Phi : G \times Q \to Q$ such that $\langle f_d, \xi_{Q\times Q}\rangle = 0$ for all $\xi \in \mathfrak{g}$. Then the discrete Lagrangian momentum map $J_{L_d}^f : Q \times Q \to \mathfrak{g}^*$ will be preserved by the discrete Lagrangian evolution map, so that $J_{L_d}^f \circ F_{L_d} = J_{L_d}^f$.*

With the above definition of the discrete Lagrangian momentum map in the presence of forcing, we see that it will be the pullback of the Hamiltonian momentum map under the forced discrete Legendre transforms, and so the discrete forced Noether's theorem can also be stated for the forced discrete Hamiltonian map $\tilde{F}_{L_d}$ with the canonical momentum map $J_H : T^*Q \to \mathfrak{g}^*$.

As in the continuous case, a similar calculation to that given above shows that the discrete symplectic form will not be preserved in the presence of forcing.

4.2.4 Exact discrete forcing

In the unforced case, we have seen that the discrete Lagrangian should approximate the continuous action over the time step. When forces are added, this must be modified so that the discrete Lagrange-d'Alembert principle (4.6) approximates the continuous expression (4.1).

Given a Lagrangian $L : TQ \to \mathbb{R}$ and a Lagrangian force $f_L : TQ \to T^*Q$, we define the ***exact forced discrete Lagrangian*** $L_d^E : Q \times Q \times \mathbb{R} \to \mathbb{R}$ and the ***exact discrete forces*** $f_d^{E+}, f_d^{E-} : Q \times Q \times \mathbb{R} \to T^*Q$ to be

$$L_d^E(q_0, q_1, h) = \int_0^h L(q(t), \dot{q}(t))dt, \tag{4.11a}$$

$$f_d^{E+}(q_0, q_1, h) = \int_0^h f_L(q(t), \dot{q}(t)) \cdot \frac{\partial q(t)}{\partial q_1} dt, \tag{4.11b}$$

$$f_d^{E-}(q_0, q_1, h) = \int_0^h f_L(q(t), \dot{q}(t)) \cdot \frac{\partial q(t)}{\partial q_0} dt, \tag{4.11c}$$

where $q : [0,h] \to Q$ is the solution of the forced Euler-Lagrange equations (4.2) for L and f_L satisfying the boundary conditions $q(0) = q_0$ and $q(h) = q_1$.

Note that this exact discrete Lagrangian is not the same as that for the unforced system with Lagrangian L, as the curves $q(t)$ are different. In other words, the exact discrete Lagrangian depends on both the continuous Lagrangian and the continuous forces, as do the discrete forces.

Given these definitions of the exact discrete quantities and the forced discrete Legendre transforms, it is easy to check that the forced version of Lemma 2.1 holds, and thus so too do forced versions of Theorems 2.14 and 2.13, showing the equivalence of the exact discrete system to the continuous systems. This is of particular interest because it shows that the variational error analysis developed in Section 3.3 can also be extended to the case of forced systems in the obvious way, and that there will be a forced version of Theorem 3.3.

Note that, if $\Phi : G \times Q \to Q$ is a symmetry of L such that $\langle f_L(q,\dot{q}), \xi_Q(q) \rangle = 0$, so the forced Noether's theorem holds, then the exact discrete forces will satisfy $\langle f_d, \xi_{Q\times Q} \rangle = 0$ and so the forced discrete Noether's theorem will also hold, as we would expect. This shows that (4.10) are the correct choice for the definition of the discrete Lagrangian momentum maps in the presence of forcing.

4.2.5 Integration of forced systems

To simulate a given forced Lagrangian or Hamiltonian system, we can choose a discrete Lagrangian and discrete forces to approximate the exact quantities given above, and then consider the resulting discrete system as an integrator for the continuous problem. We now give some simple examples of how to effect this.

Example 4.1. *The natural discrete forces for the discrete Lagrangian L_d^α given in Example 3.1 are*

$$\begin{aligned} f_d^{\alpha+}(q_0,q_1,h) &= \alpha h f_L\left((1-\alpha)q_0 + \alpha q_1, \frac{q_1-q_0}{h}\right), \\ f_d^{\alpha-}(q_0,q_1,h) &= (1-\alpha) h f_L\left((1-\alpha)q_0 + \alpha q_1, \frac{q_1-q_0}{h}\right). \end{aligned}$$

For $L = \frac{1}{2}\dot{q}^T M \dot{q} - V(q)$, the discrete Hamiltonian map is then

$$\begin{aligned} \frac{q_1-q_0}{h} &= M^{-1}\left(\alpha p_0 + (1-\alpha)p_1\right), \\ \frac{p_1-p_0}{h} &= -\nabla V\left((1-\alpha)q_0 + \alpha q_1\right) \\ &\qquad + f_H\left((1-\alpha)q_0 + \alpha q_1, \alpha p_0 + (1-\alpha)p_1\right), \end{aligned}$$

which is the same as the unforced map (3.10) with the Hamiltonian force $f_H = (\mathbb{F}L)^{-1} \circ f_L$ added to the momentum equation. For $\alpha = 1/2$ this is once again simply the midpoint rule. $\lozenge$

A particularly interesting class of Lagrangian forces $f_L : TQ \to T^*Q$ consists of those forces that

satisfy

$$\big\langle f_L(q,\dot{q}),(q,\dot{q})\big\rangle < 0,$$

for all $(q,\dot{q}) \in TQ$. Such forces are said to be (***strongly***) ***dissipative***. This terminology can be justified by computing the time evolution of the energy $E_L : TQ \to \mathbb{R}$ along a solution of the forced Euler-Lagrange equations to give

$$\begin{aligned}\frac{d}{dt}E_L(q(t),\dot{q}(t)) &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right)\cdot\dot{q} + \frac{\partial L}{\partial \dot{q}}\cdot\ddot{q} - \frac{d}{dt}L \\ &= \left(\frac{\partial L}{\partial q} + f_L\right)\cdot\dot{q} + \frac{\partial L}{\partial \dot{q}}\cdot\ddot{q} - \frac{\partial L}{\partial q}\cdot\dot{q} - \frac{\partial L}{\partial \dot{q}}\cdot\ddot{q} \\ &= f_L(q(t),\dot{q}(t))\cdot\dot{q}(t).\end{aligned}$$

We thus see that dissipative forces are those for which the energy of the system always decreases. If we only have $f_L \cdot \dot{q} \leq 0$, then the force is said to be ***weakly dissipative***.

Because the discrete Euler-Lagrange equations do not, in general, conserve energy, it is unlikely that, without some time step adaptation, there is a discrete analogue of this result.

Example 4.2. *As an example of a dissipative system, consider the movement of a unit mass particle in the plane with radial potential $V(q) = \|q\|^2(\|q\|^2-1)^2$ and forcing $f_L(q,\dot{q}) = -10^{-3}\dot{q}$. For this force we have $f_L \cdot \dot{q} = -10^{-3}\|\dot{q}\|^2 \leq 0$.*

In Figure 1.3 we plot the energy behaviour of the L_d^α method with $\alpha = 1/2$ for this system. For comparison, we also plot an extremely accurate benchmark trajectory, showing the true energy of the system, and the trajectory of the standard fourth-order Runge-Kutta method.

Observe that the variational method dissipates energy due to the discrete forces added to the Euler-Lagrange equations, but this energy dissipation is of the correct amount to accurately track the true energy. In contrast, non-conservative methods such as the Runge-Kutta integrator used here artificially dissipate energy.

These effects are of particular importance when the amount of forcing or dissipation in the system is small compared to the magnitude of the conservative dynamics and the time period of integration. For an investigation of the long time behaviour of symplectic methods applied to systems with dissipation, see Hairer and Lubich [1999]. ◊

Example 4.3 (Composition methods). *Consider a sequence of discrete Lagrangians L_d^i, discrete forces f_d^{i+}, f_d^{i-} and time step fractions γ^i for $i = 1,\ldots,s$ satisfying $\sum_{i=1}^s \gamma^i = 1$. Then we can form a composition discrete Lagrangian L_d and composition discrete forces f_d^+, f_d^- in a similar way to the procedures in Section 3.5.*

Given points q_0 and q_s, define q_i for $i = 1,\ldots,s-1$ to satisfy the forced discrete Euler-Lagrange equations (4.7) along the sequence $q_0, q_1, \ldots, q_s$. Regarding the q_i as functions of q_0 and q_1, we now

define the **composition discrete Lagrangian** *and* **composition discrete forces** *by*

$$\begin{aligned}
L_d(q_0,q_1,h) &= \sum_{i=1}^{s} L_d(q_{i-1},q_i,\gamma^i h),\\
f_d^+(q_0,q_1,h) &= f_d^{s+}(q_{s-1},q_s,\gamma^s h)\\
&\qquad + \sum_{i=1}^{s-1}\big(f_d^{i+}(q_{i-1},q_i,\gamma^i h) + f_d^{i-}(q_i,q_{i+1},\gamma^{i+1}h)\big)\cdot\frac{\partial q_i}{\partial q_1},\\
f_d^-(q_0,q_1,h) &= f_d^{1-}(q_0,q_1,\gamma^1 h)\\
&\qquad + \sum_{i=1}^{s-1}\big(f_d^{i+}(q_{i-1},q_i,\gamma^i h) + f_d^{i-}(q_i,q_{i+1},\gamma^{i+1}h)\big)\cdot\frac{\partial q_i}{\partial q_0}.
\end{aligned}$$

With these definitions it can be shown, using a similar derivation to that in Section 3.5, that the forced discrete Hamiltonian map for L_d and f_d^+, f_d^- is the composition of the individual forced discrete Hamiltonian maps, so that

$$\tilde{F}^h_{L_d} = \tilde{F}^{\gamma^s h}_{L_d^s} \circ \tilde{F}^{\gamma^{s-1} h}_{L_d^{s-1}} \circ \cdots \circ \tilde{F}^{\gamma^1 h}_{L_d^1}.$$

In forming composition methods it is often useful to use a sequence consisting of copies of a method together with its adjoint. It is thus worth noting that the adjoint of a discrete Lagrangian and discrete forces is given by

$$\begin{aligned}
L_d^*(q_0,q_1,h) &= -L_d(q_1,q_0,-h),\\
f_d^{*+}(q_0,q_1,h) &= -f_d^-(q_1,q_0,-h),\\
f_d^{*-}(q_0,q_1,h) &= -f_d^+(q_1,q_0,-h).
\end{aligned}$$

The discrete Hamiltonian map of the adjoint discrete Lagrangian and adjoint discrete forces will be the adjoint map of the original discrete Hamiltonian map. Observe that the exact discrete Lagrangian and exact discrete forces (4.11) are self-adjoint. ◊

Example 4.4 (Symplectic partitioned Runge-Kutta methods). *Recall that the discrete Lagrangian (3.28) given by*

$$L_d(q_0,q_1,h) = h\sum_{i=1}^{s} b_i L(Q_i,\dot{Q}_i)$$

generates symplectic partitioned Runge-Kutta methods. Reasonable choices of corresponding discrete

forces are

$$f_d^+(q_0, q_1, h) = h \sum_{i=1}^{s} b_i f_L(Q_i, \dot{Q}_i) \cdot \frac{\partial Q_i}{\partial q_0}, \tag{4.12a}$$

$$f_d^-(q_0, q_1, h) = h \sum_{i=1}^{s} b_i f_L(Q_i, \dot{Q}_i) \cdot \frac{\partial Q_i}{\partial q_1}, \tag{4.12b}$$

which approximate the exact forces (4.11b) and (4.11c) in the same way that L_d approximates the exact discrete Lagrangian (4.11a).

With these choices of discrete forces, it can be shown that the discrete Hamiltonian map defined by (4.9) is exactly a partitioned Runge-Kutta method for the forced Hamiltonian system (4.3). $\Diamond$

In most of the other examples of variational integrators discussed above, discrete forces can be chosen in a natural way so that the discrete Hamiltonian maps give the expected integrator for the forced Hamiltonian system. In particular, this can be done for the symplectic Newmark methods (see Kane et al. [2000]). We can also use alternative splitting-style methods to include forcing (see Kane et al. [2000] for details).

4.3 Background: Constrained systems

A particularly elegant way to study many systems is to consider them as a constrained version of some larger system. This can be appealing for both theoretical reasons and, as we shall see, also on numerical grounds. Here we will only consider so-called ***holonomic constraints***, which are constraints on the configuration manifold of a system.

More precisely, if we have a Lagrangian or Hamiltonian system with configuration manifold Q, we consider a ***constraint function*** $\phi : Q \to \mathbb{R}^d$ and constrain the dynamics to the ***constraint submanifold*** $N = \phi^{-1}(0) \subset Q$. Here we will always assume that $0 \in \mathbb{R}^d$ is a regular point of ϕ, so that N is truly a submanifold of Q [Abraham et al., 1988].

Observe that, if $i : N \to Q$ is the embedding map, then $Ti : TN \to TQ$ provides a canonical way to embed TN in TQ and we will thus regard TN as a submanifold of TQ. There is, however, no canonical way to embed the cotangent bundle T^*N in T^*Q, a fact which has important consequences for the development of constrained Hamiltonian dynamics. We will see below that, in the special case when we have a regular Lagrangian or Hamiltonian, we can use this additional structure to provide a canonical embedding.

As in other areas of mechanics, we may consider constrained systems from both the Hamiltonian and the Lagrangian viewpoint. We will concentrate on the variational approach, however, as it is this formulation which readily extends to the discrete setting. The primary tool for constrained optimization problems is the Lagrange multiplier theorem, which we recall here (see Abraham et al.

[1988] for the proof).

Theorem 4.3. *Consider a smooth manifold $\mathcal{C}$ and a function $\Phi : \mathcal{C} \to V$ mapping to some inner product space V, such that $0 \in V$ is a regular point of Φ. Set $\mathcal{D} = \Phi^{-1}(0) \subset \mathcal{C}$. Given a function $\mathfrak{G} : \mathcal{C} \to \mathbb{R}$, define $\bar{\mathfrak{G}} : \mathcal{C} \times V \to \mathbb{R}$ by $\bar{\mathfrak{G}}(q, \lambda) = \mathfrak{G}(q) - \langle \lambda, \Phi(q) \rangle$. Then the following are equivalent:*

(1). $q \in \mathcal{D}$ is an extremum of $\mathfrak{G}|_{\mathcal{D}}$;

(2). $(q, \lambda) \in \mathcal{C} \times V$ is an extremum of $\bar{\mathfrak{G}}$.

4.3.1 Constrained Lagrangian systems

Given a Lagrangian system specified by a configuration manifold Q and a Lagrangian $L : TQ \to \mathbb{R}$, consider the holonomic constraint $\phi : Q \to \mathbb{R}^d$ and the corresponding constraint submanifold $N = \phi^{-1}(0) \subset Q$. Now TN is a submanifold of TQ, and so we may restrict L to $L^N = L|_{TN}$. We are interested in the relationship of the dynamics of L^N on TN to the dynamics of L on TQ.

To consider this, we will make use of the following convenient notation. Assume that we are working on a given time interval $[0, T] \subset \mathbb{R}$, and that we have fixed endpoints $q_0, q_T \in N \subset Q$. Now set $\mathcal{C}(Q) = \mathcal{C}([0, T], Q; q_0, q_T)$ to be the space of smooth functions $q : [0, T] \to Q$ satisfying $q(0) = q_0$ and $q(T) = q_T$, and $\mathcal{C}(N)$ to be the corresponding space of curves in N. Similarly, we set $\mathcal{C}(\mathbb{R}^d) = \mathcal{C}([0, T], \mathbb{R}^d)$ to be curves $\lambda : [0, T] \to \mathbb{R}^d$ with no boundary conditions. In general $\mathcal{C}(P)$ is the space of curves from $[0, T]$ to the manifold P with the appropriate boundary conditions.

Theorem 4.4. *Given a Lagrangian system $L : TQ \to \mathbb{R}$ with holonomic constraint $\phi : Q \to \mathbb{R}^d$, set $N = \phi^{-1}(0) \subset Q$ and $L^N = L|_{TN}$. Then the following are equivalent:*

(1). $q \in \mathcal{C}(N)$ extremizes $\mathfrak{G}^N$ and hence solves the Euler-Lagrange equations for L^N;

(2). $q \in \mathcal{C}(Q)$ and $\lambda \in \mathcal{C}(\mathbb{R}^d)$ satisfy the **constrained Euler-Lagrange equations**

$$\frac{\partial L}{\partial q^i}(q(t), \dot{q}(t)) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}(q(t), \dot{q}(t))\right) = \left\langle \lambda(t), \frac{\partial \phi}{\partial q^i}(q(t)) \right\rangle, \tag{4.13a}$$

$$\phi(q(t)) = 0; \tag{4.13b}$$

(3). $(q, \lambda) \in \mathcal{C}(Q \times \mathbb{R}^d)$ extremizes $\bar{\mathfrak{G}}(q, \lambda) = \mathfrak{G}(q) - \langle \lambda, \Phi(q) \rangle$ and hence solves the Euler-Lagrange equations for the **augmented Lagrangian** *$\bar{L} : T(Q \times \mathbb{R}^d) \to \mathbb{R}$ defined by*

$$\bar{L}(q, \lambda, \dot{q}, \dot{\lambda}) = L(q, \dot{q}) - \langle \lambda, \phi(q) \rangle .$$

Proof. We make use of the Lagrange multiplier theorem, Theorem 4.3. To do so, we prepare the following definitions. The full space is $\mathcal{C} = \mathcal{C}(Q)$ and the function to be extremized is the action

$\mathfrak{G} : \mathcal{C}(Q) \to \mathbb{R}$. Take $V = \mathcal{C}(\mathbb{R}^d)$ with the L_2 inner product and define the constraint function $\Phi : \mathcal{C} \to V$ by $\Phi(q)(t) = \phi(q(t))$. Clearly $\Phi(q) = 0$, and hence $\phi(q(t)) = 0$ for all $t \in [0, T]$, if and only if $q \in \mathcal{C}(N)$. We thus obtain that the constraint submanifold is $\mathcal{D} = \Phi^{-1}(0) = \mathcal{C}(N)$.

Condition (1) simply means that $q \in \mathcal{C}(N) = \mathcal{D}$ is an extremum of the action for L^N, which is readily seen to be the standard action restricted to $\mathcal{C}(N)$. Thus $q \in \mathcal{D}$ is an extremum of $\mathfrak{G}|_{\mathcal{D}}$ and so, by the Lagrange multiplier theorem, this is equivalent to $(q, \lambda) \in \mathcal{C} \times V$ being an extremum of $\bar{\mathfrak{G}}(q, \lambda) = \mathfrak{G}(q) - \langle \lambda, \Phi(q) \rangle$.

Now $\mathcal{C} \times V = \mathcal{C}(Q) \times \mathcal{C}(\mathbb{R}^d)$ and so it can be identified with $\mathcal{C}(Q \times \mathbb{R}^d)$. Furthermore, we see that $\bar{\mathfrak{G}} : \mathcal{C}(Q \times \mathbb{R}^d) \to \mathbb{R}$ is

$$\begin{aligned}
\bar{\mathfrak{G}}(q, \lambda) &= \mathfrak{G}(q) - \langle \lambda, \Phi(q) \rangle \\
&= \int_0^T L(q(t), \dot{q}(t)) dt - \int_0^T \langle \lambda(t), \Phi(q)(t) \rangle \, dt \\
&= \int_0^T \left[L(q(t), \dot{q}(t)) - \langle \lambda(t), \phi(q(t)) \rangle \right] dt,
\end{aligned}$$

which is simply the action for the augmented Lagrangian $\bar{L}(q, \lambda, \dot{q}, \dot{\lambda}) = L(q, \dot{q}) - \langle \lambda, \phi(q) \rangle$. As $(q, \lambda) \in \mathcal{C}(Q \times \mathbb{R}^d)$ must extremize this action, we see that it is a solution of the Euler-Lagrange equations for $\bar{L}$, which is statement (3).

Finally, we extremize $\bar{\mathfrak{G}}$ by solving $\mathbf{d}\bar{\mathfrak{G}} = 0$ to obtain the Euler-Lagrange equations. The standard integration by parts argument gives (4.13a) for variations with respect to q, and variations with respect to λ imply (4.13b), and thus we have equivalence to statement (2). □

If $i : N \to Q$ is the embedding, then by differentiating $L^N = L \circ Ti$ with respect to $\dot{q}$ we see that

$$\frac{\partial L^N}{\partial \dot{q}}(v_q) \cdot w_q = \frac{\partial L}{\partial \dot{q}}\big(Ti(v_q)\big) \cdot Ti \cdot w_q, \tag{4.14}$$

which means that if L is regular, then so is L^N and shows that the following diagram commutes.

$$\begin{array}{ccc}
TQ|_N & \xrightarrow{\mathbb{F}L} & T^*Q|_N \\
{\scriptstyle Ti}\uparrow & & \downarrow{\scriptstyle T^*i} \\
TN & \xrightarrow[\mathbb{F}L^N]{} & T^*N
\end{array} \tag{4.15}$$

Using this together with the fact that $\pi_Q \circ Ti = i \circ \pi_N$ for the projections $\pi_Q : TQ \to Q$ and

$\pi_N : TN \to N$, we compute the pullback of the Lagrange one-form Θ_L on TQ to be

$$\begin{aligned}\big((Ti)^*\Theta_L\big)(v_q)\cdot\delta v_q &= \big\langle \mathbb{F}L\big(Ti(v_q)\big), T\pi_Q \circ T(Ti)\cdot\delta v_q\big\rangle \\ &= \big\langle \mathbb{F}L\big(Ti(v_q)\big), Ti\circ T\pi_N\cdot\delta v_q\big\rangle \\ &= \big\langle \mathbb{F}L^N(v_q), T\pi_N\cdot\delta v_q\big\rangle,\end{aligned}$$

and thus we see that $(Ti)^*\Theta_L = \Theta_{L^N}$, and so

$$(Ti)^*\Omega_L = \Omega_{L^N}. \tag{4.16}$$

Using the projection $T^*i : T^*Q \to T^*N$ we can reinterpret statement (2) of Theorem 4.4. Observe that the span of the $\nabla\phi^i$, $i = 1,\dots,d$ is exactly the null space of T^*i, and so (4.13) is equivalent to

$$(T^*i)_{q(t)}\left[\frac{\partial L}{\partial q}(q(t),\dot q(t)) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot q}(q(t),\dot q(t))\right)\right] = 0. \tag{4.17}$$

The above relationships hold for any Lagrangian L, irrespective of regularity. Also note that, although there is a canonical projection $T^*i : T^*Q \to T^*N$, there is no corresponding canonical embedding of T^*N into T^*Q. We will see below that when L is regular we can use the Legendre transform to define such an embedding.

4.3.2 Constrained Hamiltonian systems: Augmented approach

One can consider the Hamiltonian formulation of constrained systems by either working on the augmented space $T^*(Q\times\mathbb{R}^d)$, or working directly on T^*N, which gives the Dirac theory of constraints. We consider the former option first.

Given a Hamiltonian $H : T^*Q \to \mathbb{R}$, we define the ***augmented Hamiltonian*** to be

$$\bar H(q,\lambda,p,\pi) = H(q,p) + \langle\lambda,\phi(q)\rangle\,,$$

where π is the conjugate variable to λ. We now consider the ***primary constraint set*** $\Pi \subset T^*(Q\times\mathbb{R}^d)$ defined by $\pi = 0$. Pulling Ω back to Π gives the degenerate two-form Ω^Π, and the augmented Hamiltonian vector field $\bar X_{\bar H}$ is defined by

$$\mathbf{i}_{\bar X_{\bar H}}\Omega^\Pi = \mathbf{d}\bar H,$$

which in coordinates is the set of ***constrained Hamilton's equations***

$$\bar{X}_{q^i}(q,\lambda,p,\pi) = \frac{\partial H}{\partial p_i}, \tag{4.18a}$$

$$\bar{X}_{p_i}(q,\lambda,p,\pi) = -\frac{\partial H}{\partial q^i} - \left\langle \lambda, \frac{\partial \phi}{\partial q^i}(q) \right\rangle, \tag{4.18b}$$

$$\phi(q) = 0, \tag{4.18c}$$

where there is no λ equation owing to the degeneracy of Ω^Π. Note that for nonregular H these equations will not, in general, uniquely define the vector field $\bar{X}_{\bar{H}}$.

Consider now a regular Lagrangian L and its corresponding Hamiltonian H. Observe that the augmented Lagrangian $\bar{L}$ is degenerate, owing to the lack of dependence on $\dot{\lambda}$, and that the primary constraint manifold Π is exactly the image of $\mathbb{F}\bar{L}$. The augmented Hamiltonian and Lagrangian satisfy the equation $\bar{H} \circ \mathbb{F}\bar{L} = E_{\bar{L}}$, but this does not uniquely specify $\bar{H}$ since $\mathbb{F}\bar{L}$ need not be invertible. Nonetheless, it is simple to check that the constrained Hamilton's equations given above are equivalent to the constrained Euler-Lagrange equations (4.13) when we neglect the π component.

4.3.3 Constrained Hamiltonian systems: Dirac theory

As an alternative to working on the augmented space $T^*(Q \times \mathbb{R}^d)$, we can directly compare the dynamics of the constrained system on T^*N with those on T^*Q. The general form of this is the ***Dirac theory of constraints*** [Marsden and Ratiu, 1999], but here we use only the simple case of holonomic constraints on cotangent bundles.

The main problem with this approach is that there is no canonical way to embed T^*N within T^*Q. For now we will assume that we have an embedding $\eta : T^*N \to T^*Q$ such that $\pi_Q \circ \eta = i \circ \pi_N$ and $\eta^*\Omega = \Omega^N$, where Ω and Ω^N are the canonical two-forms on T^*Q and T^*N respectively, and we will see below how to construct η given a regular Hamiltonian or Lagrangian.

Given a Hamiltonian $H : T^*Q \to \mathbb{R}$, we define $H^N : T^*N \to \mathbb{R}$ by $H^N = H \circ \eta$. The constrained Hamiltonian vector field $X_{H^N} : T^*N \to T(T^*N)$ is then defined by

$$\mathbf{i}_{X_{H^N}} \Omega^N = \mathbf{d}H^N.$$

Taking $\pi_\Omega : T(T^*Q) \to T(T^*N)$ to be the projection operator determined by using Ω to define the orthogonal complement of $T\eta \cdot T(T^*N) \subset T(T^*Q)$, leads us to the following simple relationship between the Hamiltonian vector field X_H and the constrained vector field X_{H^N}.

Theorem 4.5. *Consider a Hamiltonian system $H : T^*Q \to \mathbb{R}$ and the corresponding constrained*

system $H^N : T^*N \to \mathbb{R}$ *as defined above. Then*

$$X_{H^N} = \pi_\Omega \cdot X_H \circ \eta.$$

Proof. We have that $\eta^*\Omega = \Omega^N$. Take an arbitrary $V^N \in T(T^*N)$ and compute

$$\begin{aligned}
\mathbf{i}_{(\pi_\Omega \cdot X_H \circ \eta)}\Omega^N \cdot V^N &= \Omega(T\eta \cdot \pi_\Omega \cdot X_H, T\eta \cdot V^N) \\
&= \Omega(X_H, T\eta \cdot V^N) \\
&= \mathbf{d}H \cdot T\eta \cdot V^N \\
&= \mathbf{d}H^N \cdot V^N \\
&= \mathbf{i}_{X_{H^N}}\Omega^N \cdot V^N,
\end{aligned}$$

where we used the fact that $(Id - T\eta \cdot \pi_\Omega) \cdot X_H$ is Ω-orthogonal to the set $T\eta \cdot T(T^*N)$. Finally, the fact that Ω^N is nondegenerate gives the desired equivalence. □

4.3.4 Legendre transforms

Until this point we have assumed that we are using any symplectic embedding $\eta : T^*N \to T^*Q$ covering the embedding $i : N \to Q$. We now consider a hyperregular Hamiltonian H and the corresponding hyperregular Lagrangian L. Recall that ***hyperregularity*** of H, for example, means that $\mathbb{F}H$ is not only a local diffeomorphism (equivalent to regularity), but is a *global* diffeomorphism. Of course, if we only have regularity, these constructions may be done locally. We will show that, using this additional structure, there is a canonical way to construct η.

To do this, begin from either a hyperregular Lagrangian L or a hyperregular Hamiltonian H, and construct the corresponding L or H, which is necessarily hyperregular as well and has $\mathbb{F}L = (\mathbb{F}H)^{-1}$. This implies that L^N and H^N are also hyperregular.

We now define $\eta : T^*N \to T^*Q$ by requiring that the following diagram commutes, where $i : N \to Q$ is the embedding as before.

$$\begin{array}{ccc}
TQ|_N & \xrightarrow{\mathbb{F}L} & T^*Q|_N \\
{\scriptstyle Ti}\uparrow & & \uparrow{\scriptstyle \eta} \\
TN & \xrightarrow[\mathbb{F}L^N]{} & T^*N
\end{array} \tag{4.19}$$

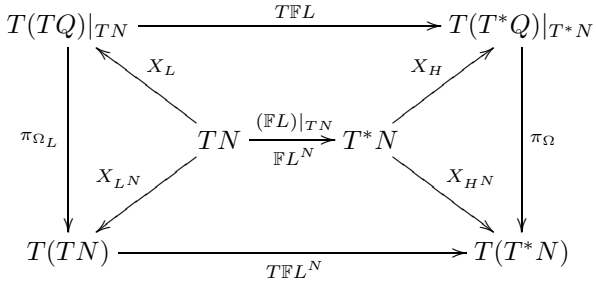
Clearly $\pi_Q \circ \eta = i \circ \pi_N$, and from (4.16) we see that $\eta^*\Omega = \Omega^N$, and so η gives a symplectic embedding of T^*N in T^*Q. Note that, although T_qN is a linear subset of T_qQ, the map η is in general not linear and so T_q^*N is not a linear subspace of T_q^*Q. It is true, however, that $T_{p_q}(T^*N)$ is a linear subspace of $T_{p_q}(T^*Q)$.

Regarding T^*N as a submanifold of T^*Q by means of η, we have the natural embedding $T\eta : T(T^*N) \to T(T^*Q)$ and so we can regard X_{H^N} as a vector field on $\eta(T^*N)$. Using canonical coordinates (q^i, p_i) on T^*Q we can derive a simple coordinate representation of this vector field:

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p}, \\ \dot{p} &= -\frac{\partial H}{\partial q} - \lambda^T \nabla \phi(q), \\ \phi(q) &= 0. \end{aligned}$$

These equations are clearly equivalent to (4.18) above if we neglect the π variable there.

Consider the projection operator $\pi_{\Omega_L} : T(TQ) \to T(TN)$ defined by the Ω_L-orthogonal complement to $T(TN)$ regarded as a subspace of $T(TQ)$ by the map TTi. As $\Omega_L = (\mathbb{F}L)^*\Omega$, elements of $T(T^*Q)$ which are Ω-orthogonal pull back under $\mathbb{F}L$ to elements of $T(TQ)$ which are Ω_L-orthogonal. It follows that $T\mathbb{F}L^n \circ \pi_{\Omega_L} = \pi_\Omega \circ T\mathbb{F}L$. In addition, observe that, as both the constrained and unconstrained systems are regular, we obtain $X_L = (\mathbb{F}L)^* X_H$ and $X_{L^N} = (\mathbb{F}L^N)^* X_{H^N}$. Combining this with the statement of Theorem 4.5 and regarding TN and T^*N as submanifolds of TQ and T^*Q, respectively, gives the following commutative diagram.



(4.20)

This establishes that $X_{L^N} = \pi_{\Omega_L} \circ X_L \circ Ti$, which is the Lagrangian analogue of Theorem 4.5. Note that this only holds for regular Lagrangians, whereas the Hamiltonian result does not require regularity.

A special case of hyperregular systems is when we have a Riemannian metric $\langle\langle \cdot, \cdot \rangle\rangle$ on Q and the Lagrangian is of the form

$$L(v_q) = \frac{1}{2} \langle\langle v_q, v_q \rangle\rangle - V \circ \pi_Q(v_q)$$

for a potential function $V : Q \to \mathbb{R}$. Computing the Legendre transform gives

$$\mathbb{F}L(v_q) \cdot w_q = \langle\langle v_q, w_q \rangle\rangle = v_q^T M(q) w_q,$$

where we introduce the symmetric positive definite mass matrix $M(q)$ as the coordinate representation of the metric. In coordinates, the Legendre transform is thus $p = M(q)\dot{q}$, and we see that the

Legendre transform is linear in $\dot{q}$ and so $\eta(T^*_qN)$ is a linear subspace of T^*_qQ at each $q \in N$. Note that the constrained subspaces can be expressed as

$$TN = \{(q, \dot{q}) \in TQ \mid \phi(q) = 0 \text{ and } \nabla\phi \cdot \dot{q} = 0\}, \tag{4.21}$$

$$\eta(T^*N) = \{(q, p) \in T^*Q \mid \phi(q) = 0 \text{ and } \nabla\phi \cdot M^{-1}(q)\dot{q} = 0\}. \tag{4.22}$$

We define the projection map $\mathbb{P} : T^*Q|_N \to \eta(T^*N)$ by $\mathbb{P} = \eta \circ T^*i$, and as it must satisfy $\mathbb{P}(\nabla\phi^m) = 0$ for each $m = 1, \ldots, d$ we can calculate the coordinate expression to be

$$\mathbb{P} = I - (\nabla\phi)^T[(\nabla\phi)M^{-1}(\nabla\phi)^T]^{-1}(\nabla\phi)M^{-1}, \tag{4.23}$$

where I is the $n \times n$ identity matrix and $\nabla\phi$ is the $d \times n$ matrix $[\nabla\phi(q)]_{mi} = \frac{\partial\phi^m}{\partial q^i}$, and all quantities are evaluated at $q \in N$.

Another way to derive this expression is to define an induced Riemannian metric on T^*Q by $\langle\langle p_q, r_q \rangle\rangle = \langle p_q, \mathbb{F}H(r_q)\rangle$, which has coordinate expression $p^T M^{-1}(q)r$. The projection $\mathbb{P}$ is then the projection onto the orthogonal subspace to the span of $\{\nabla\phi^m\}$ in the inner product given by this metric.

In this case, note that $\mathbb{P} : TQ|_N \to TN$, and so $T\mathbb{P} : T(TQ|_N) \to T(TN)$. However, observe that

$$T(TQ|_N) = \{w \in T(TQ) \mid T\pi_Q(w) \in TN\},$$

and as X_L is a second-order vector field, it satisfies $T\pi_Q \circ X_L = id$, and so we have that $X_L(v_q) \in T(TQ|_N)$ for all $v_q \in TN$. In particular, we can now show that, on the intersection of their domains, $\pi_{\Omega_L} = T\mathbb{P}$, which gives an explicit expression for the Lagrangian projection operator. This development is closely related to the expression of forces of constraint in terms of the second fundamental form (see Marsden and Ratiu [1999], Section 8.4).

4.3.5 Conservation properties

As we have seen above, the constrained systems on TN and T^*N defined by $L^N = L \circ Ti$ and $H^N = H \circ \eta$, respectively, are standard Lagrangian or Hamiltonian systems and so have the usual conservation properties.

In particular, the constrained Lagrangian system $L^N : TN \to \mathbb{R}$ will have a flow map that preserves the symplectic two-form $\Omega_{L^N} = (Ti)^*\Omega_L$, and the constrained Hamiltonian system $H^N : T^*N \to \mathbb{R}$ preserves the canonical two-form $\Omega^N = \eta^*\Omega$ on T^*N. For (hyper)regular systems, the Lagrangian and Hamiltonian two-forms are related by the Legendre transforms on both the constrained and unconstrained levels, so that $\Omega_L = (\mathbb{F}L)^*\Omega$ and $\Omega_{L^N} = (\mathbb{F}L^N)^*\Omega^N$.

Suppose that we have a group action $\Phi : G \times Q \to Q$ that leaves N invariant, that is, there is a restricted action $\Phi^N : G \times N \to N$ satisfying $i \circ \Phi^N = \Phi \circ i$. It is now a simple matter to check that the infinitesimal generators are related by

$$\begin{aligned}
\xi_Q \circ i &= Ti \circ \xi_N, \\
\xi_{TQ} \circ Ti &= T(Ti) \circ \xi_{TN}, \\
\xi_{T^*Q} \circ \eta &= T\eta \circ \xi_{T^*N},
\end{aligned}$$

and so the momentum maps satisfy

$$\begin{aligned}
J_{L^N} &= J_L \circ Ti, \\
J_{H^N} &= J_H \circ \eta.
\end{aligned}$$

Since Noether's theorem holds for both the constrained and unconstrained systems, the above relationship shows that essentially the same momentum map is preserved on both levels. Note that if the group action does not leave the constraint submanifold N invariant, however, then in general it is not possible to define J_{L^N} or J_{H^N} and there will be no constrained Noether's theorems.

4.4 Discrete variational mechanics with constraints

We now consider a discrete Lagrangian system $L_d : Q \times Q \to \mathbb{R}$ with the holonomic constraint $\phi : Q \to \mathbb{R}^d$ and corresponding constraint submanifold $N = \phi^{-1}(0) \subset Q$. As in the continuous case, the fact that $N \times N$ is naturally a submanifold of $Q \times Q$ means that we can restrict the discrete Lagrangian to $L_d^N = L_d|_{N \times N}$ to obtain a discrete Lagrangian system on $N \times N$. More precisely, we define the embedding $i^{N \times N} : N \times N \to Q \times Q$ by $i^{N \times N}(q_0, q_1) = (i(q_0), i(q_1))$.

To relate the dynamics of L_d^N to that of L_d, it is useful to introduce the notation for discrete trajectories corresponding to that used in the continuous case. Given times $\{0, h, 2h, \ldots, Nh = T\}$ and endpoints $q_0, q_T \in N$ we set $\mathbb{C}_d(Q) = \mathbb{C}_d(\{0, h, 2h, \ldots, Nh\}, Q; q_0, q_T)$ to be the set of discrete trajectories $q_d : \{0, h, 2h, \ldots, Nh\} \to Q$ satisfying $q_d(0) = q_0$ and $q_d(Nh) = q_T$, and $\mathbb{C}_d(N)$ to be the corresponding set of discrete trajectories in N.

Similarly, we denote by $\mathcal{C}_d(\mathbb{R}^d) = \mathcal{C}_d(\{h, 2h, \ldots, (N-1)h\}, \mathbb{R}^d)$ the set of maps $\lambda_d : \{h, 2h, \ldots, (N-1)h\} \to \mathbb{R}^d$ with no boundary conditions. We will see below why we do not include the boundary points 0 and Nh. In general, $\mathcal{C}_d(P)$ is the space of maps from $\{0, h, 2h, \ldots, Nh\}$ to the manifold P, and we identify such maps with their images, and write $q_d = \{q_k\}_{k=0}^N$ for $k = 0, 1, 2, \ldots, N$, and similarly for $\lambda_d = \{\lambda_k\}_{k=0}^N$.

4.4.1 Constrained discrete variational principle

As we have do not use vector fields to define the dynamics in the discrete case, and so cannot project such objects onto the constraint manifold, we turn instead to constraining the variational principle. The following theorem gives the result of this procedure.

Theorem 4.6. *Given a discrete Lagrangian system $L_d : Q \times Q \to \mathbb{R}$ with holonomic constraint $\phi : Q \to \mathbb{R}^d$, set $N = \phi^{-1}(0) \subset Q$ and $L_d^N = L_d|_{N\times N}$. Then the following are equivalent:*

(1). $q_d = \{q_k\}_{k=0}^N \in \mathcal{C}_d(N)$ extremizes $\mathfrak{G}_d^N = \mathfrak{G}_d\big|_{N\times N}$ and hence solves the discrete Euler-Lagrange equations for L_d^N;

(2). $q_d = \{q_k\}_{k=0}^N \in \mathcal{C}_d(Q)$ and $\lambda_d = \{\lambda_k\}_{k=1}^{N-1} \in \mathcal{C}_d(\mathbb{R}^d)$ satisfy the **constrained discrete Euler-Lagrange equations**

$$D_2L_d(q_{k-1}, q_k) + D_1L_d(q_k, q_{k+1}) = \langle \lambda_k, \nabla\phi(q_k)\rangle\,, \tag{4.24a}$$

$$\phi(q_k) = 0; \tag{4.24b}$$

(3). $(q_d, \lambda_d) = \{(q_k, \lambda_k)\}_{k=0}^N \in \mathcal{C}_d(Q \times \mathbb{R}^d)$ extremizes $\bar{\mathfrak{G}}_d(q_d, \lambda_d) = \mathfrak{G}_d(q_d) - \langle \lambda_d, \Phi_d(q_d)\rangle_{l_2}$ and hence solves the discrete Euler-Lagrange equations for either of the **augmented discrete Lagrangians** *$\bar{L}_d^+, \bar{L}_d^- : (Q \times \mathbb{R}^d) \times (Q \times \mathbb{R}^d) \to \mathbb{R}$ defined by*

$$\bar{L}_d^+(q_k, \lambda_k, q_{k+1}, \lambda_{k+1}) = L_d(q_k, q_{k+1}) - \langle \lambda_{k+1}, \phi(q_{k+1})\rangle\,,$$
$$\bar{L}_d^-(q_k, \lambda_k, q_{k+1}, \lambda_{k+1}) = L_d(q_k, q_{k+1}) - \langle \lambda_k, \phi(q_k)\rangle\,.$$

Proof. The proof of Theorem 4.4 in the continuous case can be almost directly applied in the discrete case.

We take the full space to be $\mathcal{C}_d = \mathcal{C}_d(Q)$ and the function we are extremizing is the discrete action $\mathfrak{G}_d : \mathcal{C}_d(Q) \to \mathbb{R}$. The constraint is specified by setting $V_d = \mathcal{C}_d(\mathbb{R}^d)$ with the l_2 inner product, and defining the constraint function $\Phi_d : \mathcal{C}_d \to V_d$ by $\Phi_d(q_d)(kh) = \phi(q_d(kh)) = \phi(q_k)$. Thus $q_d \in \mathcal{C}_d(N)$ if and only if $\phi(q_k) = 0$ for all k, and hence if and only if $\Phi_d(q_d) = 0$. The constraint submanifold is therefore $\mathcal{D}_d = \Phi_d^{-1}(0) = \mathcal{C}_d(N)$.

As in the continuous case, statement (1) means that $q_d \in \mathcal{C}_d(N) = \mathcal{D}_d$ is an extremum of the action for L_d^N, which is the full action restricted to $\mathcal{C}_d(N)$. From the Lagrange multiplier theorem (Theorem 4.3), $q_d \in \mathcal{D}_d$ being an extremum of $\mathfrak{G}_d|_{\mathcal{D}_d}$ is equivalent to $(q_d, \lambda_d) \in \mathcal{C}_d \times V_d$ being an

extremum of $\bar{\mathfrak{G}}_d(q_d, \lambda_d) = \mathfrak{G}_d(q_d) - \langle \lambda_d, \Phi_d(q_d) \rangle$. Computing, this gives

$$\begin{aligned}\bar{\mathfrak{G}}_d(q_d, \lambda_d) &= \mathfrak{G}_d(q_d) - \langle \lambda_d, \Phi_d(q_d) \rangle \\ &= \sum_{k=0}^{N-1} L_d(q_k, q_{k+1}) - \sum_{k=1}^{N-1} \langle \lambda_d(kh), \Phi_d(q_d)(kh) \rangle \,.\end{aligned}$$

Extremizing this function with respect to q_d now gives (4.24a), and extremizing with respect to λ_d recovers (4.24b). We therefore have equivalence to statement (2).

As we only extremize with respect to the internal points, and hold the boundary terms fixed, we may extend $\mathcal{C}_d(\mathbb{R}^d)$ to include λ_0 and λ_N. We now identify $\mathcal{C}_d \times V_d = \mathcal{C}_d(Q) \times \mathcal{C}_d(\mathbb{R}^d)$ with the space $\mathcal{C}_d(Q \times \mathbb{R}^d) = \mathcal{C}_d(\{0, h, 2h, \ldots, Nh\}, Q \times \mathbb{R}^d)$, and group the terms in the above expression for $\bar{\mathfrak{G}}_d$ to give two alternative functions $\bar{\mathfrak{G}}_d^+, \bar{\mathfrak{G}}_d^- : \mathcal{C}_d(Q \times \mathbb{R}^d) \to \mathbb{R}$ defined by

$$\begin{aligned}\bar{\mathfrak{G}}_d^+ &= \sum_{k=0}^{N-1} \left[L_d(q_k, q_{k+1}) - \langle \lambda_{k+1}, \phi(q_{k+1}) \rangle \right], \\ \bar{\mathfrak{G}}_d^- &= \sum_{k=0}^{N-1} \left[L_d(q_k, q_{k+1}) - \langle \lambda_k, \phi(q_k) \rangle \right],\end{aligned}$$

which have the same extrema as $\bar{\mathfrak{G}}_d$ when the boundary terms are held fixed. Identifying the terms in the summations as the augmented discrete Lagrangians $\bar{L}_d^+$ and $\bar{L}_d^-$, respectively, gives equivalence to statement (3). □

Note that in Theorem 4.6 one can actually take any convex combination of $\bar{L}_d^+$ and $\bar{L}_d^-$, although this will not substantially alter the result.

We may also use the projection operator $T^*i : T^*Q|_N \to T^*N$ to act on statement (2) of Theorem 4.6, showing that (4.24) is equivalent to

$$(T^*i)_{q_k} \left[D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}) \right] = 0. \tag{4.25}$$

This is the counterpart of the continuous equation (4.17).

4.4.2 Augmented Hamiltonian viewpoint

Just as in the continuous case, one can either work on the augmented space $T^*(Q \times \mathbb{R}^d)$ or directly on the constrained space T^*N.

The problem with trying to form the augmented discrete Hamiltonian maps $\bar{L}_d^\pm$ is the same as in this continuous case, namely the fact that the augmented discrete Lagrangians $\bar{L}_d^\pm$ are necessarily degenerate. Nonetheless, we will *define* the discrete Hamiltonian map $\tilde{F}_{\bar{L}_d^-} : (q_0, \lambda_0, p_0, \pi_0) \mapsto$

$(q_1, \lambda_1, p_1, \pi_1)$ by the equations

$$p_0 = -D_1 L_d(q_0, q_1) + \langle \lambda_0, \nabla\phi(q_0) \rangle \,, \tag{4.26a}$$

$$\pi_0 = \phi(q_0), \tag{4.26b}$$

$$p_1 = D_2 L_d(q_0, q_1), \tag{4.26c}$$

$$\pi_1 = 0. \tag{4.26d}$$

Restricting to the same primary constraint set $\Pi \subset T^*(Q \times \mathbb{R}^d)$ as in the continuous case, we see that these equations are the equivalent to (4.24) together with the requirement $\phi(q_k) = 0$ and hence $q_d \in \mathcal{C}_d(N)$, that is, they are equivalent to statement (2) in Theorem 4.6.

Note that the evolution of λ is not well defined, as in the continuous case, so that (4.26) do not define a map $\Pi \to \Pi$, that is, λ_0 is not a free initial condition, as it will be determined by (q_0, p_0). Note that constructing the alternative map $\tilde{F}_{\bar{L}_d^+}$ does not give a well-defined forward map in general. In fact, to map forward in time it is necessary to use $\tilde{F}_{\bar{L}_d^-}$ as defined above, while $\tilde{F}_{\bar{L}_d^+}$ can be used to map backward in time.

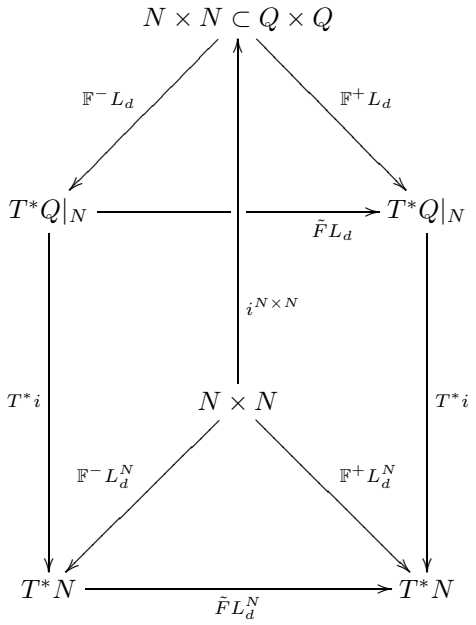
4.4.3 Direct Hamiltonian viewpoint

Alternatively, one can neglect the augmented space and directly relate T^*N and T^*Q. To do so, we differentiate $L_d^N = L_d \circ i^{N\times N}$ with respect to q_0 and q_1 to obtain the discrete equivalents of (4.14), thus establishing that the following diagrams commute.

$$\begin{array}{ccc} T^*Q|_N & \xleftarrow{\mathbb{F}^- L_d} & Q \times Q|_{\{q_0 \in N\}} \\ {\scriptstyle T^*i}\downarrow & & \uparrow{\scriptstyle i^{N\times N}} \\ T^*N & \xleftarrow[\mathbb{F}^- L_d^N]{} & N \times N \end{array} \qquad \begin{array}{ccc} Q \times Q|_{\{q_1 \in N\}} & \xrightarrow{\mathbb{F}^+ L_d} & T^*Q|_N \\ {\scriptstyle i^{N\times N}}\uparrow & & \downarrow{\scriptstyle T^*i} \\ N \times N & \xrightarrow[\mathbb{F}^+ L_d^N]{} & T^*N \end{array} \tag{4.27}$$

We will henceforth assume that L_d is regular, which means that L_d^N is also regular and that the discrete Hamiltonian maps $\tilde{F}_{L_d}$ and $\tilde{F}_{L_d^N}$ are well defined. Combining the above diagrams with the expressions $\tilde{F}_{L_d} = \mathbb{F}^+ L_d \circ (\mathbb{F}^- L_d)^{-1} : T^*Q \to T^*Q$ and $\tilde{F}_{L_d^N} = \mathbb{F}^+ L_d^N \circ (\mathbb{F}^- L_d^N)^{-1} : T^*N \to T^*N$

gives the following commutative diagram.



(4.28)

This proves the following theorem.

Theorem 4.7. *Consider a regular discrete Lagrangian system $L_d : Q \times Q \to \mathbb{R}$ and the constrained system $L_d^N : N \times N \to \mathbb{R}$ defined by $L_d^N = L_d \circ i^{N\times N}$. Then the discrete Hamiltonian map $\tilde{F}_{L_d^N} : T^*N \to T^*N$ has the following equivalent formulations:*

(1). $\tilde{F}_{L_d^N} : (q_0, p_0) \mapsto (q_1, p_1)$ for $(q_0, p_0), (q_1, p_1) \in T^*N$ satisfying

$$p_0 = -D_1 L_d^N(q_0, q_1), \tag{4.29a}$$

$$p_1 = D_2 L_d^N(q_0, q_1); \tag{4.29b}$$

(2). $\tilde{F}_{L_d^N} : (q_0, p_0) \mapsto (q_1, p_1)$ for $(q_0, p_0), (q_1, p_1) \in T^*N$ satisfying

$$p_0 = (T^*i)_{q_0}(-D_1 L_d \circ i^{N\times N}(q_0, q_1)), \tag{4.30a}$$

$$p_1 = (T^*i)_{q_1}(D_2 L_d \circ i^{N\times N}(q_0, q_1)); \tag{4.30b}$$

(3). $\tilde{F}_{L_d^N} : \eta(T^*N) \mapsto \eta(T^*N)$ for $(q_0, p_0) \in \eta(T^*N)$ and $(q_1, p_1) \in T^*Q$ satisfying

$$p_0 = \mathbb{P}_{q_0}(-D_1 L_d \circ i^{N\times N}(q_0, q_1)), \tag{4.31a}$$

$$p_1 = \mathbb{P}_{q_1}(D_2 L_d \circ i^{N\times N}(q_0, q_1)), \tag{4.31b}$$

$$\phi(q_1) = 0. \tag{4.31c}$$

Here $\eta : T^*N \to T^*Q$ is any symplectic embedding covering the identity, so that $\pi_Q \circ \eta = i \circ \pi_N$, and $\mathbb{P} : T^*Q|_N \to \eta(T^*N)$ is the map defined by $\mathbb{P} = \eta \circ T^*i$.

This theorem is the discrete analogue of Theorem 4.5, and shows how the unconstrained Hamiltonian equations are related to the constrained equations. If we further assume that η is defined by (4.19) for some regular Lagrangian L with corresponding Hamiltonian H, then we can use the fact that the null space of $\mathbb{P}$ is the span of the $\nabla\phi^m$ and introduce Lagrange multipliers to write (4.31) as

$$p_0 = -D_1 L_d(q_0, q_1) + (\lambda^{(0)})^T \nabla\phi(q_0), \tag{4.32a}$$

$$p_1 = D_2 L_d(q_0, q_1) - (\lambda^{(1)})^T \nabla\phi(q_1), \tag{4.32b}$$

$$\phi(q_1) = 0, \tag{4.32c}$$

$$\nabla\phi(q_1) \cdot \frac{\partial H}{\partial p}(q_1, p_1) = 0, \tag{4.32d}$$

defining a map from $(q_0, p_0) \in \eta(T^*N)$ to $(q_1, p_1) \in T^*Q$ which will satisfy $(q_1, p_1) \in \eta(T^*N)$. Here the arbitrary signs on the Lagrange multipliers have been chosen to correspond to the signs for discrete forces in (4.9).

Now consider the special case when Q is a Riemannian manifold with metric $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ having coordinate representation $M(q)$ and η is defined by (4.19) for a Lagrangian with kinetic energy given by the metric. As we have seen in Section 4.3.4 above, $\eta(T^*N)$ and $\mathbb{P}$ are now given explicitly by (4.22) and (4.23), respectively. Using this, we can write (4.31) as

$$p_0 = -\big(I - (\nabla\phi)^T \big[(\nabla\phi) M^{-1} (\nabla\phi)^T\big]^{-1} (\nabla\phi) M^{-1}\big) D_1 L_d(q_0, q_1), \tag{4.33a}$$

$$p_1 = \big(I - (\nabla\phi)^T \big[(\nabla\phi) M^{-1} (\nabla\phi)^T\big]^{-1} (\nabla\phi) M^{-1}\big) D_2 L_d(q_0, q_1), \tag{4.33b}$$

$$\phi(q_1) = 0, \tag{4.33c}$$

where $\nabla\phi$ and M are evaluated at q_0 or q_1 as appropriate.

4.4.4 Conservation properties

A constrained discrete Lagrangian system on $N \times N$ and an unconstrained system on $Q \times Q$ will clearly preserve the standard discrete symplectic two-forms $\Omega_{L_d^N}$ and Ω_{L_d}, respectively. Now define the projections $\pi_Q^1 : Q \times Q \to Q$ and $\pi_N^1 : N \times N \to N$ onto the first components of $Q \times Q$ and $N \times N$. Observe that $\pi_Q^1 \circ i^{N\times N} = i \circ \pi_N^1$ and, together with the left-hand diagram in (4.27), a similar calculation to that preceding equation (4.16) will now establish that $\Theta^-_{L_d^N} = (i^{N\times N})^* \Theta^-_{L_d}$. Using the same idea for $\Theta^+_{L_d}$ and taking the exterior derivative of these expressions shows that the

constrained and unconstrained discrete one- and two-forms are related by

$$\Theta^+_{L_d^N} = (i^{N\times N})^*\Theta^+_{L_d}, \qquad \Theta^-_{L_d^N} = (i^{N\times N})^*\Theta^-_{L_d}, \qquad \Omega_{L_d^N} = (i^{N\times N})^*\Omega_{L_d}.$$

Pushing all of these structures forward with the discrete Legendre transforms shows that the constrained discrete Hamiltonian map $\tilde{F}_{L_d^N}$, regarded as acting either on T^*N or $\eta(T^*N)$, preserves the canonical two-form Ω^N, while $\tilde{F}_{L_d}$ naturally preserves Ω.

If we further consider a symmetry action $\Phi : G \times Q \to Q$ which leaves N invariant, so that it covers an action $\Phi^N : G \times N \to N$, then the infinitesimal generators are related by

$$\xi_Q \circ i = Ti \circ \xi_N, \tag{4.34a}$$

$$\xi_{Q\times Q} \circ i^{N\times N} = T(i^{N\times N}) \circ \xi_{N\times N}. \tag{4.34b}$$

Using now the above relations between the constrained and unconstrained symplectic one-forms, we have that the momentum maps for the product action will be related by

$$J^+_{L_d^N} = J^+_{L_d} \circ i^{N\times N}, \tag{4.35a}$$

$$J^-_{L_d^N} = J^-_{L_d} \circ i^{N\times N}. \tag{4.35b}$$

If the group action is a symmetry of the Lagrangian, then these momentum maps are equal and Noether's theorem holds on both the constrained and unconstrained levels with this unique momentum map.

4.4.5 Constrained exact discrete Lagrangians

The exact discrete Lagrangian for a constrained system is not simply the standard exact discrete Lagrangian restricted to the constraint submanifold, as that would be the action along an unconstrained trajectory. Instead, the constrained exact discrete Lagrangian is the action of the constrained system, evaluated along the trajectory which lies on the constraint submanifold: that is,

$$L_d^{N,E}(q_0, q_1, h) = \int_0^h L^N(q_{0,1}(t), \dot{q}_{0,1}(t))dt, \tag{4.36}$$

where $q : [0, h] \to N$ is the solution of the Euler-Lagrange equations for $L^N : TN \to \mathbb{R}$ which satisfies $q(0) = q_0$ and $q(h) = q_1$. As this discrete Lagrangian is defined on $N \times N \times \mathbb{R}$, it satisfies

$$\mathbb{F}^- L_d^{N,E}(q_0, q_1, h)) = \mathbb{F}L^N(q_{0,1}(0), \dot{q}_{0,1}(0)),$$
$$\mathbb{F}^+ L_d^{N,E}(q_0, q_1, h)) = \mathbb{F}L^N(q_{0,1}(h), \dot{q}_{0,1}(h)).$$

We would like, however, to define a function on $Q \times Q \times \mathbb{R}$ whose restriction to $N \times N \times \mathbb{R}$ would give $L_d^{N,E}$. Without introducing additional structure, however, there is no canonical way to do so. Indeed, let $L_d^{Q,E} : Q \times Q \times \mathbb{R} \to \mathbb{R}$ be any smooth extension of $L_d^{N,E}$. Then from (4.15), (4.27) and the above relations we have immediately that

$$(T^*i)_{q_0}\big(\mathbb{F}^- L_d^{Q,E}(q_0, q_1, h)\big) = (T^*i)_{q_0}\big(\mathbb{F}L(q_{0,1}(0), \dot{q}_{0,1}(0))\big),$$
$$(T^*i)_{q_1}\big(\mathbb{F}^+ L_d^{Q,E}(q_0, q_1, h)\big) = (T^*i)_{q_1}\big(\mathbb{F}L(q_{0,1}(h), \dot{q}_{0,1}(h))\big),$$

which is a constrained version of Lemma 2.1. The equivalence of the discrete and continuous systems now follows as in Section 2.5.

Note that this means that the order of accuracy of a discrete Lagrangian constrained to $N \times N$ will not, in general, be the same as the order of accuracy on $Q \times Q$: that is, if $L_d : Q \times Q \times \mathbb{R} \to \mathbb{R}$ approximates the action on Q to some particular order, then the restriction $L_d^N = L_d|_{N \times N}$ will typically approximate the action of constrained solutions in N to some different order. Indeed, to derive high-order discrete Lagrangians for a constrained system, it is necessary to take account of the constraints in defining L_d^N, since a high-order L_d will typically restrict to only a first- or second-order L_d^N.

4.5 Constrained variational integrators

In this section we consider implementing the integration of a mechanical system with constraints. First we review standard geometric methods, and then we turn to variational integrators.

4.5.1 Constrained geometric integration

There are a number of standard approaches to the numerical integration of constrained mechanical systems. These include working in local coordinates on the constraint submanifold (for example, see Bobenko and Suris [1999b] in the case of Lie groups, or Leimkuhler and Reich [1994]), solving a modified system on the containing space which has the constraint submanifold as a stable invariant set (for example, see Leimkuhler and Reich [1994]), and methods based in the containing space which explicitly enforce the constraints. Constrained mechanical systems are particular examples of differential algebraic systems, and many of the techniques for the numerical integration of such systems can also be applied in the mechanical setting (see Hairer and Wanner [1996] and Ascher and Petzold [1998]).

Unless the system under consideration has a particularly simple structure, working in local coordinates on the constraint submanifold suffers from a number of problems, including the fact that changing charts during the integration is not smooth, which breaks many of the nice properties

of geometric integrators. In addition, local coordinate computations can be very expensive, and the equations can be very complicated, making the integrator difficult to code. For all of these reasons, it is often preferable to use integration techniques based on the containing space.

There are a number of different approaches to this, with representative samples being Gonzalez [1999], Seiler [1999, 1998b,a], McLachlan and Scovel [1995] and Brasey and Hairer [1993]. For a good overview of this area see Hairer [2000].

4.5.2 Variational integrators for constrained systems

Here we consider a constrained discrete Lagrangian system as an integrator for a continuous system. Given a continuous system $L : TQ \to \mathbb{R}$ and a constraint submanifold $N \subset Q$ defined by $N = \phi^{-1}(0)$ for some $\phi : Q \to \mathbb{R}^d$, we would like a discrete Lagrangian $L_d : Q \times Q \times \mathbb{R} \to \mathbb{R}$ so that its restriction to $N \times N \times \mathbb{R}$ approximates the exact constrained discrete Lagrangian (4.36). The order of this approximation is related to the order of the resulting integrator.

Given such an L_d, we can now use any of the equivalent formulations of the constrained Euler-Lagrange equations from Section 4.4.3 to obtain an integrator. As in the unconstrained case, we can regard such an integrator as defined on the product $N \times N$ or on the corresponding cotangent bundle, although the latter interpretation is typically simpler for implementation purposes.

To be explicit, we will henceforth assume that the given continuous system is regular, so that we have equivalent Lagrangian and Hamiltonian representations, and that the containing manifold Q is linear, so that it is isomorphic to $\mathbb{R}^n$. We will use (4.32) to define the constrained discrete Hamiltonian map $\tilde{F}_{L_d^N}$ regarded as mapping $\eta(T^*N)$ to $\eta(T^*N)$, where we recall that $\eta(T^*N)$ is the embedding of T^*N in T^*Q defined by

$$\eta(T^*N) = \Big\{(q,p) \in T^*Q \;\mid\; \phi(q) = 0 \text{ and } \nabla\phi(q) \cdot \frac{\partial H}{\partial p}(q,p) = 0\Big\}. \tag{4.37}$$

As we are now treating the discrete Lagrangian as the approximation to the exact system, it will be dependent upon a time step h and thus have the form $L_d(q_0, q_1, h)$. Given this, we may rescale the Lagrange multipliers in (4.32) by h so that the constraint terms appear in the same way as discrete forces, allowing them to be interpreted as discrete forces of constraint. This gives

$$p_0 = -D_1 L_d(q_0, q_1) + h(\lambda^{(0)})^T \nabla\phi(q_0), \tag{4.38a}$$

$$p_1 = D_2 L_d(q_0, q_1) - h(\lambda^{(1)})^T \nabla\phi(q_1), \tag{4.38b}$$

$$\phi(q_1) = 0, \tag{4.38c}$$

$$\nabla\phi(q_1) \cdot \frac{\partial H}{\partial p}(q_1, p_1) = 0. \tag{4.38d}$$

To use these equations as an integrator, we must take an initial condition $(q_0, p_0) \in \eta(T^*N)$, so that q_0 and p_0 satisfy the conditions given by (4.37). The $2n + 2d$ system (4.38) must then be solved implicitly to find (q_1, p_1) and the accompanying Lagrange multipliers. Iterating this process gives the integrated trajectory.

Although this is generally the simplest way to implement a variational integrator, note that if the Lagrangian has a special form, such as being composed of kinetic and potential terms, then we could also use one of the other equivalent expressions of the discrete Hamiltonian map given previously. Alternatively, we could also choose to work directly on $N \times N$ and to use (4.24) as an integrator mapping each pair (q_k, q_{k+1}) to (q_{k+1}, q_{k+2}).

Using the above theory, we recall that any such methods will always be symplectic, and if the discrete Lagrangian inherits the symmetries of the continuous system, then the integrator will also conserve the corresponding momentum maps.

To implement a constrained variational integrator, it is of course necessary to choose a particular discrete Lagrangian. We give below a number of ways in which this can be done and we explicitly evaluate the defining equations (4.38) in several cases.

4.5.3 Low-order methods

Given a low-order discrete Lagrangian, such as L_d^α given in Example 3.1, one can simply restrict it to $N \times N$ to obtain an integrator for the constrained system. As N will generally not be convex, the points $(1-\alpha)q_0 + \alpha q_1$ will not be in N when q_0 and q_1 are. If the Lagrangian on N is the restriction of a smooth Lagrangian on Q, then this will not matter for sufficiently small step sizes.

For a Lagrangian which is not defined off N, or which varies quickly compared to the step size, it is important to only evaluate L and its derivatives on N. Perhaps the simplest examples of such methods are given by L_d^0 and L_d^1, which give constrained versions of the symplectic Euler methods.

4.5.4 SHAKE and RATTLE

As we saw in Section 3.6.2, the Verlet algorithm is the discrete Lagrangian map $F_{L_d} : Q \times Q \to Q \times Q$ generated by the discrete Lagrangian

$$L_d(q_0, q_1, h) = \frac{1}{2} h L\left(q_0, \frac{q_1 - q_0}{h}\right) + \frac{1}{2} h L\left(q_1, \frac{q_1 - q_0}{h}\right), \tag{4.39}$$

where we assume that the continuous system has the form $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M \dot{q} - V(q)$. To form a constrained version of this method, we can simply restrict L_d to $N \times N$ and calculate the constrained

discrete Euler-Lagrange equations (4.24). These give

$$M\left(\frac{q_{k+1}-2q_k+q_{k-1}}{h}\right)+h\nabla V(q_k)+(\lambda_k)^T\nabla\phi(q_k)=0,$$
$$\phi(q_{k+1})=0,$$

which is known as the SHAKE algorithm. This was first proposed by Ryckaert et al. [1977] as a constrained version of Verlet.

A constrained version of the velocity Verlet integrator, RATTLE, was given by Anderson [1983]. This was later shown by Leimkuhler and Skeel [1994] to be a symplectic integrator on T^*N. In fact, RATTLE *is simply the constrained discrete Hamiltonian map* $\tilde{F}_{L_d^N}:T^*N\to T^*N$ *associated to the discrete Lagrangian* (4.39). To see this, we calculate the coordinate expressions of (4.38) with $L(q,\dot{q})=\frac{1}{2}\dot{q}^TM\dot{q}-V(q)$ to give

$$\begin{aligned}
p_k &= M\left(\frac{q_{k+1}-q_k}{h}\right)+\frac{1}{2}h\nabla V(q_k)+(\lambda_k^{(0)})^T\nabla\phi(q_k),\\
p_{k+1} &= M\left(\frac{q_{k+1}-q_k}{h}\right)-\frac{1}{2}h\nabla V(q_{k+1})+(\lambda_k^{(1)})^T\nabla\phi(q_{k+1}),\\
0 &= \phi(q_{k+1}),\\
0 &= \nabla\phi(q_{k+1})M^{-1}p_{k+1}.
\end{aligned}$$

Now we subtract the first equation from the second and solve the first equation for q_{k+1} to obtain

$$\begin{aligned}
q_{k+1} &= q_k+hM^{-1}p_k+\frac{1}{2}h^2M^{-1}(-\nabla V(q_k))+\frac{1}{2}h^2M^{-1}(\lambda_k^{(0)})^T\nabla\phi(q_k),\\
p_{k+1} &= p_k+h\left(\frac{-\nabla V(q_k)-\nabla V(q_{k+1})}{2}\right)\\
&\qquad +h\left(\frac{(\lambda_k^{(0)})^T\nabla\phi(q_k)+(\lambda_k^{(1)})^T\nabla\phi(q_{k+1})}{2}\right),\\
0 &= \phi(q_{k+1}),\\
0 &= \nabla\phi(q_{k+1})M^{-1}p_{k+1},
\end{aligned}$$

where we are assuming

$$(\nabla\phi)_{ij}(q)=\frac{\partial\phi^i}{\partial q^j}$$

and where we have scaled $\lambda_k^{(0)}$ and $\lambda_k^{(1)}$ by $-\frac{1}{2}$. This is exactly the RATTLE method.

This integrator is also the 2-stage member of the Lobatto IIIA-IIIB family [Jay, 1996, 1999], which is discussed further below.

To summarize, the integrators known as Verlet, velocity Verlet, SHAKE and RATTLE are all

derived from the discrete Lagrangian (4.39). Verlet is the discrete Lagrangian map $F_{L_d} : Q \times Q \to Q \times Q$, velocity Verlet is the discrete Hamiltonian map $\tilde{F}_{L_d} : T^*Q \to T^*Q$, SHAKE is the constrained discrete Lagrangian map $F_{L_d^N} : N \times N \to N \times N$, and RATTLE is the constrained discrete Hamiltonian map $\tilde{F}_{L_d^N} : T^*N \to T^*N$.

Thus, the variational formulation shows the natural connection between these methods, and proves in a unified way that they all conserve both the symplectic structure and quadratic momentum maps, as linear symmetries of V will be inherited by L_d.

4.5.5 Composition methods

To construct high-order integrators for a constrained system, a simple low-order constraint-preserving method can be used in a composition rule, as in Section 3.5 [Reich, 1996]. This approach has the advantage that the resulting method will inherit properties such as symplecticity from the base method, and will necessarily preserve the constraint.

Composing discrete Lagrangians extends directly to constrained systems. Given discrete Lagrangians L_d^i and time step fractions γ^i for $i = 1, \ldots, s$, we can use any of the three interpretations of the composition L_d from Section 3.5. For the multiple steps method or the single step, multiple substeps method, the correct constraint to impose is that all the points q_k^i lie on the constraint submanifold. This implies that the single step constrained composition discrete Lagrangian should be defined as

$$L_d(q_k, q_{k+1}, h) = \operatorname*{ext}_{q_k^i \in N} L_d(q_k, q_k^i, q_{k+1}, h),$$

which denotes the extreme value of the multipoint discrete Lagrangian over the set of interior points in the constraint submanifold N. The constrained discrete Hamiltonian map for this L_d will then be the composition of the constrained discrete Hamiltonian maps of the component L_d^i.

When composing non-self-adjoint methods, it is common to use a sequence including both the methods themselves and their adjoints. For this reason, it is worth noting that the adjoint of a constrained discrete Lagrangian is equal to the constrained version of the adjoint, that is, $(L_d^*)^N = (L_d^N)^*$. Furthermore, the associated constrained discrete Hamiltonian maps are adjoint as integrators.

4.5.6 Constrained symplectic partitioned Runge-Kutta methods

For a Hamiltonian system $H : T^*Q \to \mathbb{R}$ with holonomic constraint $\phi : Q \to \mathbb{R}^d$, a ***constrained partitioned Runge-Kutta method*** is a map $T^*N \to T^*N$ specified by $(q_0, p_0) \mapsto (q_1, p_1)$ where

$$q_1 = q_0 + h\sum_{j=1}^{s} b_j \dot{Q}_j, \qquad p_1 = p_0 + h\sum_{j=1}^{s} \tilde{b}_j \dot{P}_j, \tag{4.40a}$$

$$Q_i = q_0 + h\sum_{j=1}^{s} a_{ij} \dot{Q}_j, \qquad P_i = p_0 + h\sum_{j=1}^{s} \tilde{a}_{ij} \dot{P}_j, \qquad i = 1, \ldots, s, \tag{4.40b}$$

$$\dot{Q}_i = \frac{\partial H}{\partial p}(Q_i, P_i), \qquad \dot{P}_i = -\frac{\partial H}{\partial q}(Q_i, P_i) - \Lambda_i^T \nabla\phi(Q_i), \qquad i = 1, \ldots, s, \tag{4.40c}$$

$$0 = \phi(Q_i), \qquad 0 = \nabla\phi(q_1) \cdot \frac{\partial H}{\partial p}(q_1, p_1), \qquad i = 1, \ldots, s. \tag{4.40d}$$

In addition, it is necessary to place some restrictions on the coefficients to ensure that these equations do in fact define a map on T^*N. We begin by imposing the requirement (3.27) of symplecticity to give

$$b_i \tilde{a}_{ij} + \tilde{b}_j a_{ji} = b_i \tilde{b}_j, \qquad i, j = 1, \ldots, s,$$
$$b_i = \tilde{b}_i, \qquad i = 1, \ldots, s.$$

We also require that the method be ***stiffly accurate***: that is, $a_{si} = b_i$ for $i = 1, \ldots, s$. This means that $q_1 = Q_s$, and hence $q_1 \in N$. Further requiring that $b_i \neq 0$ for $i = 1, \ldots, s$ implies that $\tilde{a}_{is} = 0$ for each $i = 1, \ldots, s$.

To ensure that the system is not over-determined, we set $a_{1i} = 0$ for $i = 1, \ldots, s$ and so obtain $q_0 = Q_1$. Requiring that $b_i \neq 0$ for $i = 1, \ldots, s$ now implies that $\tilde{a}_{i1} = \tilde{b}_i$ for $i = 1, \ldots, s$. Given that we start from $(q_0, p_0) \in T^*N$ we thus have that $\phi(Q_1) = \phi(q_0) = 0$ is immediately satisfied.

With these restrictions, (4.40) is a system of $s(4n + d) + 2n$ equations for the same number of unknowns, defining a map $\eta(T^*N) \to \eta(T^*N)$. It can be shown [Jay, 1996] that this is a well-defined symplectic map on T^*N. Such methods are a particular example of the SPARK methods of Jay [1999], and the subset of these methods which are explicit have been analysed for constrained systems by Reich [1997].

To see how such constrained symplectic partitioned Runge-Kutta methods can be derived variationally, we proceed in a similar fashion to the unconstrained case in Section 3.6.5. Given $(q_0, q_1) \in Q \times Q$, we implicitly define $\bar{p}_0$, $\bar{p}_1$, $\bar{Q}_i$, $\bar{P}_i$, $\dot{\bar{Q}}_i$, $\dot{\bar{P}}_i$ for $i = 1, \ldots, s$, and $\bar{\Lambda}_i$ for $i = 2, \ldots, (s-1)$ by the equations

$$q_1 = q_0 + h\sum_{j=1}^{s} b_j \dot{\bar{Q}}_j, \qquad \bar{p}_1 = \bar{p}_0 + h\sum_{j=1}^{s} \tilde{b}_j \dot{\bar{P}}_j, \tag{4.41a}$$

$$\bar{Q}_i = q_0 + h\sum_{j=1}^{s} a_{ij}\dot{\bar{Q}}_j, \qquad \bar{P}_i = \bar{p}_0 + h\sum_{j=1}^{s} \tilde{a}_{ij}\dot{\bar{P}}_j, \qquad i = 1,\dots,s, \tag{4.41b}$$

$$\dot{\bar{Q}}_i = \frac{\partial H}{\partial p}(\bar{Q}_i, \bar{P}_i), \qquad i = 1,\dots,s, \tag{4.41c}$$

$$\dot{\bar{P}}_i = -\frac{\partial H}{\partial q}(\bar{Q}_i, \bar{P}_i) - \bar{\Lambda}_i^T \nabla\phi(\bar{Q}_i), \qquad 0 = \phi(\bar{Q}_i), \qquad i = 2,\dots,(s-1), \tag{4.41d}$$

$$\dot{\bar{P}}_1 = -\frac{\partial H}{\partial q}(\bar{Q}_1, \bar{P}_1), \qquad \dot{\bar{P}}_s = -\frac{\partial H}{\partial q}(\bar{Q}_s, \bar{P}_s). \tag{4.41e}$$

This is a system of $4sn + (s-2)d$ equations in the same number of variables and the restrictions on the coefficients ensure that it will have a solution for sufficiently small h.

This subset of the equations (4.40) was chosen from the fact that $\bar{Q}_1 = q_0$ and $\bar{Q}_s = q_1$, so it is necessary to relax the constraints on these two points. Having done so, the same number of Lagrange multipliers must also then be disregarded. Given these definitions of the various quantities in terms of q_0 and q_1 we define the discrete Lagrangian $L_d : Q \times Q \times \mathbb{R} \to \mathbb{R}$ by

$$L_d(q_0, q_1, h) = h\sum_{i=1}^{s} b_i L(\bar{Q}_i, \dot{\bar{Q}}_i), \tag{4.42}$$

where we assume that the coefficients satisfy all of the previous requirements. For a given continuous system (L or H) this is not the same as the corresponding expression (3.28) in the unconstrained case, as the equations defining $\bar{Q}_i$ and $\dot{\bar{Q}}_i$ have been modified here to take account of the constraints. We now show that the constrained discrete Hamiltonian map corresponding to (4.42) is indeed the constrained symplectic partitioned Runge-Kutta method.

Theorem 4.8. *The constrained discrete Hamiltonian map for the discrete Lagrangian (4.42) is exactly the integrator defined by the constrained symplectic partitioned Runge-Kutta equations (4.40).*

Proof. Differentiating $\phi(\bar{Q}_i) = 0$ for $i = 2,\dots,s-1$ gives

$$\nabla\phi(\bar{Q}_i) \cdot \frac{\partial \bar{Q}_i}{\partial q_0} = 0, \qquad i = 2,\dots,s-1,$$

and using this together with the definitions (4.41) and the same argument as in Theorem 3.6 shows that

$$\frac{\partial L_d}{\partial q_0} = -\bar{p}_0, \qquad \frac{\partial L_d}{\partial q_1} = \bar{p}_1.$$

We now consider a given initial condition $(q_0, p_0) \in T^*N$ and recall that the discrete Hamiltonian map will give $(q_1, p_1) \in T^*N$ which satisfy (4.38). To see the relation of this mapping to the

symplectic partitioned Runge-Kutta map, we make the following change of variables:

$$\begin{aligned}
Q_i &= \bar{Q}_i, & P_i &= \bar{P}_i, & i &= 1, \ldots, s, \\
\dot{Q}_i &= \dot{\bar{Q}}_i, & & & i &= 1, \ldots, s, \\
\Lambda_i &= \bar{\Lambda}_i, & \dot{P}_i &= \dot{\bar{P}}_i, & i &= 2, \ldots, s-1, \\
\tilde{b}_1 \Lambda_1 &= \lambda^{(0)}, & \dot{P}_1 &= \dot{\bar{P}}_1 - \Lambda_1^T \nabla \phi(Q_1), & & \\
\tilde{b}_s \Lambda_s &= \lambda^{(1)}, & \dot{P}_s &= \dot{\bar{P}}_s - \Lambda_s^T \nabla \phi(Q_s). & &
\end{aligned}$$

Recalling that the coefficients are such that $Q_1 = q_0$ and $Q_s = q_1$, we now see that (4.38c) and (4.38d), together with the restrictions (4.41d) on $\bar{Q}_i$, give the conditions (4.40d) on the non-overbar quantities.

Furthermore, (4.38a) and (4.38b) give

$$\begin{aligned}
\bar{p}_0 &= p_0 - h\tilde{b}_1 \Lambda_1^T \nabla \phi(Q_1), \\
\bar{p}_1 &= p_1 + h\tilde{b}_s \Lambda_s^T \nabla \phi(Q_s).
\end{aligned}$$

Substituting these definitions into the equations (4.41) and using the fact that $\tilde{a}_{is} = 0$ and $\tilde{a}_{i1} = \tilde{b}_i$ for $i = 1, \ldots, s$ now shows that the non-overbar quantities satisfy (4.40a), (4.40b) and (4.40c). We thus have that the discrete Hamiltonian map $(q_0, p_0) \mapsto (q_1, p_1)$ on $\eta(T^*N)$ is identical to the constrained symplectic Runge-Kutta map. □

4.5.7 Constrained Galerkin methods

With the insight gained from the definition of the constrained exact discrete Lagrangian (4.36) it is simple to extend the Galerkin discrete Lagrangians of Section 3.6.6 to include holonomic constraints.

In the particular example of polynomial trajectory approximations and numerical quadrature, the definition (3.30) of the Galerkin discrete Lagrangian should be modified to

$$L_d(q_0, q_1, h) = \operatorname*{ext}_{\substack{q \in \mathcal{C}^s([0,h],Q) \\ \phi(q(c_i h))=0}} \mathfrak{G}^s(q), \tag{4.43}$$

where $\phi : Q \to \mathbb{R}$ is the constraint function. This constrains the intermediate trajectories to intersect the constraint submanifold at each quadrature point. For such methods it is typically reasonable to require that $c_0 = 0$ and $c_s = 1$, so that the endpoints q_0 and q_1 also satisfy the constraint.

Evaluating the constrained discrete Euler-Lagrange equations for (4.43) shows that the associated discrete Hamiltonian map is a constrained symplectic partitioned Runge-Kutta method, in the sense of the preceding section and of Jay [1999]. In particular, choosing the quadrature rule to be Lobatto

quadrature results in the constrained Lobatto IIIA-IIIB method of Jay [1999].

4.6 Background: Forced and constrained systems

We now consider Lagrangian and Hamiltonian systems with *both* external forcing and holonomic constraints. The formulations and equations for such systems are straightforward combinations of the material in the preceding sections for systems with only forces or only constraints. For this reason, we will simply state the results without proof.

As before, we assume that we have a system on the unconstrained configuration manifold Q, and a holonomic constraint function $\phi : Q \to \mathbb{R}^d$ so that the constraint manifold is $N = \phi^{-1}(0) \subset Q$. The inclusion map is denoted $i : N \to Q$, and we have the natural lifts $Ti : TN \to TQ$ and $T^*i : T^*Q \to T^*N$.

4.6.1 Lagrangian systems

Given a Lagrangian force $f_L : TQ \to T^*Q$, we restrict it to $f_L^N = T^*i \circ f_L \circ Ti : TN \to T^*N$, which is then a Lagrangian force on TN. Taking the Lagrange-d'Alembert principle and restricting to the space of constrained curves gives the following theorem.

Theorem 4.9. *Given a Lagrangian system $L : TQ \to \mathbb{R}$ with Lagrangian force $f_L : TQ \to T^*Q$ and holonomic constraint $\phi : Q \to \mathbb{R}^d$, set $N = \phi^{-1}(0) \subset Q$, $f_L^N = T^*i \circ f_L \circ Ti$, and $L^N = L|_{TN}$. Then the following are equivalent:*

(1). $q \in \mathcal{C}(N)$ satisfies the Lagrange-d'Alembert principle for L^N and f_L^N and hence solves the forced Euler-Lagrange equations;

(2). $q \in \mathcal{C}(Q)$ and $\lambda \in \mathcal{C}(\mathbb{R}^d)$ satisfy the **forced constrained Euler-Lagrange equations**

$$\frac{\partial L}{\partial q^i}(q(t), \dot{q}(t)) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}(q(t), \dot{q}(t))\right) + f_L(q(t), \dot{q}(t)) = \left\langle \lambda(t), \frac{\partial \phi}{\partial q^i}(q(t)) \right\rangle, \tag{4.44a}$$

$$\phi(q(t)) = 0; \tag{4.44b}$$

(3). $(q, \lambda) \in \mathcal{C}(Q \times \mathbb{R}^d)$ satisfies the Lagrange-d'Alembert principle, and hence solves the forced Euler-Lagrange equations, for $\bar{L} : T^*(Q \times \mathbb{R}^d) \to \mathbb{R}$ and $\bar{f}_L : T(Q \times \mathbb{R}^d) \to T^*(Q \times \mathbb{R}^d)$ defined by

$$\bar{L}(q, \lambda, \dot{q}, \dot{\lambda}) = L(q, \dot{q}) - \langle \lambda, \phi(q) \rangle,$$
$$\bar{f}_L(q, \lambda, \dot{q}, \dot{\lambda}) = \pi_Q^* \circ f_L(q, \dot{q}),$$

where $\pi_Q : Q \times \mathbb{R}^d \to Q$ is the projection.

One can also project (4.44a) with $T^*i : T^*Q \to T^*N$ to obtain a system without λ, as in Section 4.3.

Observe that in the forced constrained Euler-Lagrange equations (4.44) the forcing and Lagrange multiplier terms enter in same way. For this reason, the Lagrange multiplier term is sometimes referred to as the ***forces of constraint***, and we can regard it as being a force which is constructed exactly so that the solution is kept on the constraint submanifold N.

4.6.2 Hamiltonian systems

Following the development of the unforced constrained case, we can move to the Hamiltonian framework by either taking the Legendre transform of the degenerate augmented system, or by working directly on T^*N.

The former approach takes a Hamiltonian force $f_H : T^*Q \to T^*Q$ and forms the augmented Hamiltonian force $\bar{f}_H : T^*(Q \times \mathbb{R}^d) \to T^*(Q \times \mathbb{R}^d)$ by $\bar{f}_H(q, \lambda, p, \pi) = \pi_Q^* \circ f_H(q, p)$. The forced constrained Hamiltonian vector field $\bar{X}_H$ on the primary constraint set Π is defined by

$$\mathbf{i}_{\bar{X}_H} \Omega^\Pi = \mathbf{d}\bar{H} - \bar{f}'_H,$$

where $\bar{H}$ and Ω^Π are as before, and $\bar{f}'_H$ is the horizontal one-form on $T^*(Q \times \mathbb{R}^d)$ corresponding to $\bar{f}_H$. In coordinates this gives the ***forced constrained Hamilton equations***

$$\begin{aligned} X_{q^i}(q, \lambda, p, \pi) &= \frac{\partial H}{\partial p^i}, \\ X_{p^i}(q, \lambda, p, \pi) &= -\frac{\partial H}{\partial q^i} + f_H(q, p) - \left\langle \lambda, \frac{\partial \phi}{\partial q^i}(q) \right\rangle, \\ \phi(q) &= 0. \end{aligned}$$

Alternatively, we can directly relate the unconstrained Hamiltonian system to the constrained system as in Section 4.3.3. To do this, we must choose a symplectic embedding $\eta : T^*N \to T^*Q$, which we will assume covers the embedding $i : N \to Q$. Given such a map, we now define the ***constrained Hamiltonian force*** $f_H^N : T^*N \to T^*N$ by $f_H^N = T^*i \circ f_H \circ \eta$ and we let $f_H^{N\prime}$ be the corresponding horizontal one-form on T^*N. We assume that all other structures are as in Section 4.3.3, so that the constrained Hamiltonian is $H^N = H \circ \eta$.

The forced constrained Hamiltonian vector field X_{H^N} and the forced unconstrained Hamiltonian

vector field X_H are now defined by

$$\mathbf{i}_{X_{H^N}}\Omega^N = \mathbf{d}H^N - f_H^{N\prime},$$
$$\mathbf{i}_{X_H}\Omega = \mathbf{d}H - f_H'.$$

Denoting the Ω-orthogonal projection to $\eta(T^*N)$ by $\pi_\Omega : T^*Q \to T^*N$, we can show that the projection of the forced unconstrained vector field is just the forced constrained vector field.

Theorem 4.10. *Consider a Hamiltonian system $H : T^*Q \to \mathbb{R}$ with forcing $f_H : T^*Q \to T^*Q$ and constraint submanifold $N \subset Q$ and let the constrained system $H^N : T^*N \to \mathbb{R}$ and $f_H^N : T^*N \to T^*N$ be defined as above. Then $X_{H^N} = \pi_\Omega \cdot X_H \circ \eta$.*

Proof. We can use essentially the same proof as for Theorem 4.5 in the unforced case. The only additional requirement is to check that the one-form $f_H^{N\prime}$ is the pullback under η of f_H', so that $f_H'(\eta(p_q)) \cdot T\eta \cdot V^N = f_H^{N\prime}(p_q) \cdot V^N$.

To see this, we recall that η covers the identity and so $\pi_Q \circ \eta = i \circ \pi_N$. Using the derivative of this expression we calculate

$$\begin{aligned} f_H^{N\prime}(p_q) \cdot V^N &= \left\langle T^*i \circ f_H \circ \eta(p_q), T\pi_N \cdot V^N \right\rangle \\ &= \left\langle f_H \circ \eta(p_q), Ti \circ T\pi_N \cdot V^N \right\rangle \\ &= \left\langle f_H \circ \eta(p_q), T\pi_Q \circ T\eta \cdot V^N \right\rangle \\ &= \big(\eta^*(f_H')\big)(p_q) \cdot V^N, \end{aligned}$$

which can then be used to modify the proof of Theorem 4.5, to obtain the desired result. □

4.6.3 Legendre transforms

Given a regular Lagrangian system and the corresponding regular Hamiltonian system, we have seen in Section 4.3.4 that the standard Legendre transforms provide a canonical way to construct a map $\eta : T^*N \to T^*Q$ and so to regard T^*N as a submanifold of T^*Q.

Furthermore, as we saw in Section 4.1.3, the forced Lagrangian and Hamiltonian vector fields are related by the standard Legendre transform, so this will hold for both the constrained and unconstrained systems. Note that our definitions of constrained Lagrangian and Hamiltonian forces commute with the Legendre transform, so that if $f_L = f_H \circ \mathbb{F}L$, then $f_L^N = f_H^N \circ \mathbb{F}L^N$. This can be seen by recalling that $\eta \circ \mathbb{F}L^N = \mathbb{F}L \circ Ti$ and using the definitions of the constrained forces.

We thus have that the constrained and unconstrained forced vector fields on both the Lagrangian and Hamiltonian sides are related by projection and Legendre transforms, which fully commute. In

particular, we can write the projected vector field on the Hamiltonian side in coordinates to give

$$\begin{aligned}\dot{q} &= \frac{\partial H}{\partial p},\\ \dot{p} &= -\frac{\partial H}{\partial q} - \lambda^T \nabla\phi(q) + f_H(q,p),\\ \phi(q) &= 0.\end{aligned}$$

In the special case when the Hamiltonian depends quadratically on p, then this projection is induced by the metric given on T^*Q by the kinetic energy, as in Section 4.3.4 above.

4.6.4 Conservation properties

Given a group action $\Phi : G \times Q \to Q$, we have seen in Section 4.4.4 that if Φ leaves N invariant, then it can be restricted to an action Φ^N on N and the infinitesimal generators of this restricted action are related by projection to the generators of the action on Q. This then shows that the momentum maps of the constrained systems are just the appropriate restrictions of the unconstrained momentum maps.

In addition, from Section 4.1.4 we know that if the Lagrangian is invariant under the group action and the forces are orthogonal to the action, then Noether's theorem will still hold. In the constrained setting, observe that we have

$$\begin{aligned}\left\langle f_L^N(v_q), \xi_N(q)\right\rangle &= \left\langle T^*i \circ f_L \circ Ti(v_q), \xi_N(q)\right\rangle\\ &= \left\langle f_L \circ Ti(v_q), Ti \cdot \xi_N(q)\right\rangle\\ &= \left\langle f_L \circ Ti(v_q), \xi_Q \circ i(q)\right\rangle,\end{aligned}$$

and so if f_L is orthogonal to ξ_Q, then the constrained force f_L^N will also be orthogonal to the constrained infinitesimal generator ξ_N. This gives us the following Noether's theorem.

Theorem 4.11 (Forced constrained Noether's theorem). *Consider a Lagrangian system $L : TQ \to \mathbb{R}$ with constraint submanifold $N \subset Q$, forcing $f_L : TQ \to T^*Q$ and a symmetry action $\Phi : G \times Q \to Q$ such that $\langle f_L(q,\dot{q}), \xi_Q(q)\rangle = 0$ for all $(q,\dot{q}) \in TQ$ and $\xi \in \mathfrak{g}$. Then the constrained Lagrangian momentum map $J_{L^N} : TN \to \mathfrak{g}^*$ will be preserved by the forced constrained Lagrangian flow.*

Of course, it is only necessary that the constrained force be orthogonal to the group action on the constraint submanifold and that the reduced action be a symmetry of the constrained Lagrangian. The above theorem simply gives sufficient conditions for this in terms of the unconstrained quantities.

4.7 Discrete variational mechanics with forces and constraints

We now combine the previous results for forced and constrained systems to consider discrete Lagrangian systems with *both* forcing and constraints. The definitions and results are the expected combinations of the special cases of only forcing or only constraints, and so we will not give detailed proofs.

4.7.1 Lagrangian viewpoint

Given discrete Lagrangian forces $f_d^+, f_d^- : Q \times Q \to T^*Q$, we form the restrictions $f_d^{N+}, f_d^{N-} : N \times N \to T^*N$ by $f_d^{N\pm} = T^*i \circ f_d^\pm \circ i^{N\times N}$, which are then discrete Lagrangian forces on N. As in the continuous Lagrangian case, we now take the discrete Lagrange-d'Alembert principle from Section 4.4 and constrain it to N, thus obtaining the following theorem.

Theorem 4.12. *Given discrete Lagrangian system $L_d : Q \times Q \to \mathbb{R}$ with discrete Lagrangian forces $f_d^+, f_d^- : Q \times Q \to T^*Q$ and holonomic constraint $\phi : Q \to \mathbb{R}^d$, set $N = \phi^{-1}(0) \subset Q$, $f_d^{N\pm} = T^*i \circ f_d^\pm \circ i^{N\times N}$, and $L_d^N = L_d|_{Q\times Q}$. Then the following are equivalent:*

(1). $q_d = \{q_k\}_{k=0}^N \in \mathcal{C}_d(N)$ satisfies the discrete Lagrange-d'Alembert principle for L_d^N, f_d^{N+} and f_d^{N-}, and hence solves the forced discrete Euler-Lagrange equations;

(2). $q_d = \{q_k\}_{k=0}^N \in \mathcal{C}_d(Q)$ and $\lambda_d = \{\lambda_k\}_{k=1}^{N-1} \in \mathcal{C}_d(\mathbb{R}^d)$ satisfy the **forced constrained discrete Euler-Lagrange equations**

$$\begin{aligned} D_2L_d(q_{k-1},q_k) + D_1L_d(q_k,q_{k+1}) & \\ + f_d^+(q_{k-1},q_k) + f_d^-(q_k,q_{k+1}) &= \langle \lambda_k, \nabla\phi(q_k)\rangle\,, \qquad (4.45a)\\ \phi(q_k) &= 0; \qquad (4.45b) \end{aligned}$$

(3). $(q_d,\lambda_d) = \{(q_k,\lambda_k)\}_{k=0}^N \in \mathcal{C}_d(Q\times\mathbb{R}^d)$ satisfies the discrete Lagrange-d'Alembert principle, and hence solves the forced discrete Euler-Lagrange equations, for either of $\bar{L}_d^+, \bar{L}_d^- : (Q\times\mathbb{R}^d)\times(Q\times\mathbb{R}^d) \to \mathbb{R}$ defined by

$$\begin{aligned} \bar{L}_d^+(q_k,\lambda_k,q_{k+1},\lambda_{k+1}) &= L_d(q_k,q_{k+1}) - \langle \lambda_{k+1}, \phi(q_{k+1})\rangle\,,\\ \bar{L}_d^-(q_k,\lambda_k,q_{k+1},\lambda_{k+1}) &= L_d(q_k,q_{k+1}) - \langle \lambda_k, \phi(q_k)\rangle\,, \end{aligned}$$

with the discrete Lagrangian forces $\bar{f}_d^+, \bar{f}_d^- : (Q\times\mathbb{R}^d)\times(Q\times\mathbb{R}^d) \to T^*(Q\times\mathbb{R}^d)$ defined by

$$\begin{aligned} \bar{f}_d^+(q_k,\lambda_k,q_{k+1},\lambda_{k+1}) &= \pi_Q^* \circ f_d^+(q_k,q_{k+1}),\\ \bar{f}_d^-(q_k,\lambda_k,q_{k+1},\lambda_{k+1}) &= \pi_Q^* \circ f_d^-(q_k,q_{k+1}), \end{aligned}$$

where $\pi_Q : Q \times \mathbb{R}^d \to Q$ is the projection.

Using the canonical projection operator $T^*i : T^*Q \to T^*N$, we can also write (4.45) without the Lagrange multipliers.

4.7.2 Discrete Hamiltonian maps

We first consider the augmented approach to constructing a discrete Hamiltonian map, despite the lack of regularity. The ***forced augmented discrete Hamiltonian map*** $\tilde{F}_{\bar{L}_d^-} : (q_0, \lambda_0, p_0, \pi_0) \mapsto (q_1, \lambda_1, p_1, \pi_1)$ is defined by the equations

$$p_0 = -D_1 L_d(q_0, q_1) - f_d^-(q_0, q_1) + \langle \lambda_0, \nabla\phi(q_0) \rangle \,, \tag{4.46a}$$

$$\pi_0 = \phi(q_0), \tag{4.46b}$$

$$p_1 = D_2 L_d(q_0, q_1) + f_d^+(q_0, q_1), \tag{4.46c}$$

$$\pi_1 = 0. \tag{4.46d}$$

Restricting to the primary constraint set $\Pi \subset T^*(Q \times \mathbb{R}^d)$ now shows that these equations are equivalent to the forced constrained discrete Euler-Lagrange equations (4.45) together with the constraint $\phi(q_k) = 0$. As before, the evolution of λ is not well defined.

Rather than considering the augmented systems, we can also directly relate the constrained and unconstrained systems. Here we must use the forced discrete Legendre transforms (4.8), which we recall are

$$\begin{aligned}
\mathbb{F}^{f+}L_d &: (q_0, q_1) \mapsto (q_1, p_1) = (q_1, D_2 L_d(q_0, q_1) + f_d^+(q_0, q_1)), \\
\mathbb{F}^{f-}L_d &: (q_0, q_1) \mapsto (q_0, p_0) = (q_0, -D_1 L_d(q_0, q_1) - f_d^-(q_0, q_1)).
\end{aligned}$$

These depend on both the discrete Lagrangian and discrete forces. From (4.27) we have the relations

$$\begin{aligned}
D_2 L_d^N &= T^*i \circ D_2 L_d \circ i^{N \times N}, \\
-D_1 L_d^N &= T^*i \circ (-D_1 L_d) \circ i^{N \times N},
\end{aligned}$$

and, combining these with the definitions of the constrained discrete forces f_d^{N+} and f_d^{N-}, we have the following commutative diagrams, where the discrete Legendre transforms are those which include the forcing.

$$\begin{array}{ccc}
T^*Q|_N & \xleftarrow{\mathbb{F}^{f-}L_d} & Q \times Q|_{\{q_0 \in N\}} \\
{\scriptstyle T^*i}\downarrow & & \uparrow{\scriptstyle i^{N\times N}} \\
T^*N & \xleftarrow[\mathbb{F}^{f-}L_d^N]{} & N \times N
\end{array}
\qquad\qquad
\begin{array}{ccc}
Q \times Q|_{\{q_1 \in N\}} & \xrightarrow{\mathbb{F}^{f+}L_d} & T^*Q|_N \\
{\scriptstyle i^{N\times N}}\uparrow & & \downarrow{\scriptstyle T^*i} \\
N \times N & \xrightarrow[\mathbb{F}^{f+}L_d^N]{} & T^*N
\end{array} \tag{4.47}$$

This is the equivalent of (4.27) in the unforced case, and using this we now have the equivalent of diagram (4.28) for the forced discrete Legendre transforms, proving the following theorem.

Theorem 4.13. *Consider a regular discrete Lagrangian system $L_d : Q \times Q \to \mathbb{R}$ with constraint submanifold $N \subset Q$ and forcing $f_d^+, f_d^- : Q \times Q \to T^*Q$. Then the forced constrained discrete Hamiltonian map $\tilde{F}_{L_d^N} : T^*N \to T^*N$ has the following equivalent formulations:*

(1). $\tilde{F}_{L_d^N} : (q_0, p_0) \mapsto (q_1, p_1)$ for $(q_0, p_0), (q_1, p_1) \in T^*N$ satisfying

$$p_0 = -D_1 L_d^N(q_0, q_1) - f_d^{N-}(q_0, q_1), \tag{4.48a}$$

$$p_1 = D_2 L_d^N(q_0, q_1) + f_d^{N+}(q_0, q_1); \tag{4.48b}$$

(2). $\tilde{F}_{L_d^N} : (q_0, p_0) \mapsto (q_1, p_1)$ for $(q_0, p_0), (q_1, p_1) \in T^*N$ satisfying

$$p_0 = (T^*i)_{q_0}\big((-D_1 L_d - f_d^-) \circ i^{N\times N}(q_0, q_1)\big), \tag{4.49a}$$

$$p_1 = (T^*i)_{q_1}\big((D_2 L_d + f_d^+) \circ i^{N\times N}(q_0, q_1)\big); \tag{4.49b}$$

(3). $\tilde{F}_{L_d^N} : \eta(T^*N) \mapsto \eta(T^*N)$ for $(q_0, p_0) \in \eta(T^*N)$ and $(q_1, p_1) \in T^*Q$ satisfying

$$p_0 = \mathbb{P}_{q_0}\big((-D_1 L_d - f_d^-) \circ i^{N\times N}(q_0, q_1)\big), \tag{4.50a}$$

$$p_1 = \mathbb{P}_{q_1}\big((D_2 L_d + f_d^+) \circ i^{N\times N}(q_0, q_1)\big), \tag{4.50b}$$

$$\phi(q_1) = 0. \tag{4.50c}$$

Here $\eta : T^*N \to T^*Q$ is any symplectic embedding covering the identity, so that $\pi_Q \circ \eta = i \circ \pi_N$, and $\mathbb{P} : T^*Q|_N \to \eta(T^*N)$ is the map defined by $\mathbb{P} = \eta \circ T^*i$.

These equations are clearly the combination of the constrained equations from Theorem 4.7 with the forced equations (4.9).

Now assume that η is constructed from the Legendre transforms of some regular Lagrangian according to (4.19). Introducing Lagrange multipliers allows us to rewrite (4.50) as

$$p_0 = -D_1 L_d(q_0, q_1) - f_d^-(q_0, q_1) + (\lambda^{(0)})^T \nabla\phi(q_0), \tag{4.51a}$$

$$p_1 = D_2 L_d(q_0, q_1) + f_d^+(q_0, q_1) - (\lambda^{(1)})^T \nabla\phi(q_1), \tag{4.51b}$$

$$\phi(q_1) = 0, \tag{4.51c}$$

$$\nabla\phi(q_1) \cdot \frac{\partial H}{\partial p}(q_1, p_1) = 0, \tag{4.51d}$$

where (q_0, p_0) are in $\eta(T^*N)$. As before, we have chosen the signs on the Lagrange multipliers to

correspond with the conventions of the discrete forces.

This form of the forced constrained discrete Hamiltonian map shows clearly that one can interpret the Lagrange multiplier terms as ***discrete forces of constraint***. That is, the additional terms due to the constraints enter the equations in exactly the same way as the forcing terms. Indeed, the constraint terms can be regarded as forces which have exactly the correct action to keep the discrete trajectory on the constraint submanifold N.

If we are working with a particular form of Lagrangian, such as one involving a quadratic kinetic energy, then we can explicitly write the projection form of the discrete Hamiltonian map as was done in Section 4.4.3.

4.7.3 Exact forced constrained discrete Lagrangian

Given a Lagrangian system with forces and constraints, we can combine the ideas from Sections 4.2.4 and 4.4.5 to define the appropriate exact discrete Lagrangian and exact discrete forces.

Begin by considering the constrained system $L^N : TN \to \mathbb{R}$ with constrained force $f_L^N : TN \to T^*N$. Recall that the exact forced discrete Lagrangian $L_d^{N,E} : N \times N \times \mathbb{R}$ is the action (4.11a) along a solution of the forced Euler-Lagrange equations, and that the exact discrete forces $f_d^{N,E+}, f_d^{N,E-} : N \times N \times \mathbb{R} \to T^*N$ are the integrals of the forces (4.11b), (4.11c) along the variations of such a solution.

Having constructed these functions on $N \times N \times \mathbb{R}$, we take any smooth extension to functions $L_d^{Q,E} : Q \times Q \times \mathbb{R}$ and $f_d^{Q,E+}, f_d^{Q,E-} : Q \times Q \times \mathbb{R} \to \mathbb{R}$, as in Section 4.4.5. The same argument as used there now shows that

$$\begin{aligned}(T^*i)_{q_0}\big(\mathbb{F}^{f-}L_d^{Q,E}(q_0,q_1,h)\big) &= (T^*i)_{q_0}\big(\mathbb{F}L(q_{0,1}(0),\dot{q}_{0,1}(0))\big),\\ (T^*i)_{q_1}\big(\mathbb{F}^{f+}L_d^{Q,E}(q_0,q_1,h)\big) &= (T^*i)_{q_1}\big(\mathbb{F}L(q_{0,1}(h),\dot{q}_{0,1}(h))\big),\end{aligned}$$

for all $q_0, q_1 \in N$ and the corresponding solutions $q : [0,h] \to N$ of the forced constrained Euler-Lagrange equations.

Using the above definitions, it is clear that to derive high-order discrete Lagrangians and discrete forces in the presence of constraints, both the discrete Lagrangian and the discrete forces will have to depend upon the continuous Lagrangian, the continuous forces and also the constraints. We will see examples of this below.

4.7.4 Noether's theorem

Consider a group action $\Phi : G \times Q \to Q$ and assume that it leaves N invariant, so that it restricts to $\Phi^N : G \times N \to N$. In the presence of forcing we saw in Section 4.2.3 that it is necessary to use the

forced Legendre transforms to define the discrete momentum maps by (4.10). For the unconstrained system this gives

$$J_{L_d}^{f+}(q_0,q_1)\cdot\xi = \left\langle \mathbb{F}^{f+}L_d(q_0,q_1), \xi_Q(q_1)\right\rangle, \tag{4.52a}$$

$$J_{L_d}^{f-}(q_0,q_1)\cdot\xi = \left\langle \mathbb{F}^{f-}L_d(q_0,q_1), \xi_Q(q_1)\right\rangle, \tag{4.52b}$$

while the constrained forced momentum maps are

$$J_{L_d^N}^{f+}(q_0,q_1)\cdot\xi = \left\langle \mathbb{F}^{f+}L_d^N(q_0,q_1), \xi_N(q_1)\right\rangle, \tag{4.53a}$$

$$J_{L_d^N}^{f-}(q_0,q_1)\cdot\xi = \left\langle \mathbb{F}^{f-}L_d^N(q_0,q_1), \xi_N(q_1)\right\rangle. \tag{4.53b}$$

Recalling that the forced discrete Legendre transforms satisfy (4.47), we can use the relations (4.34) between the constrained and unconstrained infinitesimal generators to show that

$$J_{L_d^N}^{f+} = J_{L_d}^{f+}\circ i^{N\times N}, \tag{4.54a}$$

$$J_{L_d^N}^{f-} = J_{L_d}^{f-}\circ i^{N\times N}, \tag{4.54b}$$

which is the forced equivalent of (4.35). If the group action is a symmetry of the discrete Lagrangian then these momentum maps will be equal. In general Noether's theorem does not hold in the presence of forcing, except in the special case when the forces are orthogonal to the group action. We will now see how this occurs in the presence of constraints.

Recall that, given discrete forces f_d^+ and f_d^-, we can construct a one-form f_d on $Q\times Q$ by (4.5), which gives

$$f_d^N(q_0,q_1)\cdot(\delta q_0,\delta q_1) = f_d^{N+}(q_0,q_1)\cdot\delta q_1 + f_d^{N-}(q_0,q_1)\cdot\delta q_0,$$

$$f_d(q_0,q_1)\cdot(\delta q_0,\delta q_1) = f_d^{+}(q_0,q_1)\cdot\delta q_1 + f_d^{-}(q_0,q_1)\cdot\delta q_0,$$

and so we have the relation $f_d^N = T^*(i^{N\times N})\circ f_d\circ i^{N\times N}$. Using this, we compute

$$\begin{aligned}\left\langle f_d^N(q_0,q_1), \xi_{N\times N}(q_0,q_1)\right\rangle &= \left\langle T^*(i^{N\times N})\circ f_d\circ i^{N\times N}(q_0,q_1), \xi_{N\times N}(q_0,q_1)\right\rangle \\ &= \left\langle f_d\circ i^{N\times N}(q_0,q_1), T(i^{N\times N})\circ\xi_{N\times N}(q_0,q_1)\right\rangle \\ &= \left\langle f_d\circ i^{N\times N}(q_0,q_1), \xi_{Q\times Q}\circ i^{N\times N}(q_0,q_1)\right\rangle,\end{aligned}$$

where we have used the fact that $\xi_{Q\times Q}\circ i^{N\times N} = T(i^{N\times N})\circ\xi_{N\times N}$. This shows that if f_d is orthogonal to $\xi_{Q\times Q}$, so that $\langle f_d, \xi_{Q\times Q}\rangle = 0$, then f_d^N will be orthogonal to $\xi_{N\times N}$. We thus have a Noether's theorem in this case.

Theorem 4.14 (Discrete forced constrained Noether's theorem). *Consider a discrete Lagrangian system $L_d : Q \times Q \to \mathbb{R}$ with constraint submanifold $N \subset Q$, discrete forces $f_d^+, f_d^- : Q \times Q \to T^*Q$ and a symmetry action $\Phi : G \times Q \to Q$ such that $\langle f_d, \xi_{Q\times Q}\rangle = 0$ for all $\xi \in \mathfrak{g}$. Then the constrained Lagrangian momentum map $J^f_{L_d^N} : N \times N \to \mathfrak{g}^*$ is preserved by the forced constrained discrete Hamiltonian map.*

As in the continuous case with forcing and constraints, this only provides a sufficient condition as it is enough to just have orthogonality and invariance on N.

4.7.5 Variational integrators with forces and constraints

Consider a Lagrangian system $L : TQ \to \mathbb{R}$ with a constraint submanifold $N \subset Q$ specified by $N = \phi^{-1}(0)$ for some $\phi : Q \to \mathbb{R}^d$ and a Lagrangian force $f_L : TQ \to T^*Q$. We would now like to construct a discrete Lagrangian $L_d : Q \times Q \to \mathbb{R}$ and discrete forces $f_d^+, f_d^- : Q \times Q \to T^*Q$ which approximate an extension of the exact discrete Lagrangian and exact forces. The discrete Hamiltonian map will then be an integrator for the continuous system.

We will assume here that the Lagrangian is regular, so that it has an equivalent Hamiltonian formulation, and also that Q is linear and isomorphic to $\mathbb{R}^n$. Regularity of the Lagrangian also provides a canonical embedding $\eta : T^*N \to T^*Q$, and we will use the Lagrange multiplier formulation (4.51) of the forced constrained Hamiltonian map. As in Section 4.5.2, we will rescale the Lagrange multipliers by the time step to give

$$p_0 = -D_1 L_d(q_0, q_1) - f_d^-(q_0, q_1) + h(\lambda^{(0)})^T \nabla\phi(q_0), \tag{4.55a}$$

$$p_1 = D_2 L_d(q_0, q_1) + f_d^+(q_0, q_1) - h(\lambda^{(1)})^T \nabla\phi(q_1), \tag{4.55b}$$

$$\phi(q_1) = 0, \tag{4.55c}$$

$$\nabla\phi(q_1) \cdot \frac{\partial H}{\partial p}(q_1, p_1) = 0, \tag{4.55d}$$

where the initial condition (q_0, p_0) is in $\eta(T^*N)$, and we solve over $(q_1, p_1) \in T^*Q$. The last two equations ensure that the solution (q_1, p_1) will also lie in $\eta(T^*N)$. Of course, we could also use one of the alternative formulations from Theorem 4.13 or we could use the forced constrained discrete Euler-Lagrange equations (4.45) and work directly on $N \times N$.

To construct discrete Lagrangians and discrete forces we can use any of the techniques discussed previously. Here we give a few examples.

Example 4.5 (Low-order methods). *For a low-order discrete Lagrangian and discrete forces, such as the L_d^α and $f_d^{\alpha,\pm}$ from Example 4.1, we can simply restrict them to $N \times N$, as in Section 4.5.3. This yields a simple method that remains on the constraint manifold and includes the forcing.* $\lozenge$

Example 4.6 (Composition methods). *As we have seen in several examples already, composition methods provide a particularly elegant method to construct high-order methods from a given low-order integrator. In the case of systems with both forcing and constraints, the appropriate composed discrete forces and discrete Lagrangians are given by the combination of the definitions for the forced and constrained cases.* ◊

Example 4.7 (Symplectic partitioned Runge-Kutta methods). *Combining the definitions of the discrete forces (4.12) with the constrained formulation of the discrete Lagrangian (4.42), we arrive at discrete forces and a discrete Lagrangian for which the discrete Hamiltonian map is a constrained symplectic partitioned Runge-Kutta method with forcing.* ◊

Chapter 5

Multisymplectic continuum mechanics

5.1 Multisymplectic continuum mechanics

The basic objects for a material picture of continuum mechanics are a ***reference configuration*** $\mathcal{B} \subset \mathbb{R}^n$ of the body, a ***time interval*** $[0, T] \subset \mathbb{R}$ and an ***ambient space*** $\mathcal{S} = \mathbb{R}^m$. We then consider the ***configuration map*** $\varphi_t : \mathcal{B} \to \mathcal{S}$ which defines the particle placement or configuration at each time t.

We will now develop this theory in a multisymplectic formulation, both for continuous space-time and for discretizations. We will construct the AVI (Asynchronous Variational Integrator) methods as a particular case of the discrete multisymplectic formalism. The material below is formulated intrinsically in Marsden et al. [2001], but here we will restrict ourselves to Euclidean spaces. For more on multisymplectic mechanics and multisymplectic discretizations, see Marsden et al. [1998], Gotay et al. [1997], and Bridges and Reich [2001a]. The differential geometry notation used here follows Abraham et al. [1988]. We only consider first-order theories, in which both the Lagrangian and the constraints depend only on *first* derivatives of the fields. For higher-order formulations see Kouranbaeva and Shkoller [2000].

5.1.1 Configuration geometry

Base space. The ***base space*** $\mathcal{X} = \mathbb{R} \times \mathbb{R}^3$ is defined to be space-time. Coordinates on $\mathcal{X}$ are $(X^0 \equiv t, X^1, \ldots, X^n)$, and we will sometimes write (t, X) to distinguish the time and space coordinates. Lowercase greek letters range over $0, 1, \ldots, n$, so that X^μ is all base space coordinates. Alternatively, lowercase roman letters i, j, k range over $1, 2, \ldots, n$, and we write $t = X^0$ for time, so $(X^\mu) = (t, X^i)$. We will abuse the notation and use the symbol X to denote points in both the base space and the reference configuration $\mathcal{B}$, explicitly distinguishing when there is the possibility of confusion.

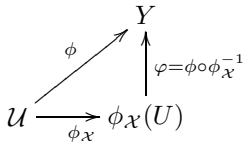
We introduce the ***parameter space*** $\mathcal{U} = [0, T] \times \mathcal{B}$. This will allow us to consider variations of

the base space variables. Coordinates on $\mathcal{U}$ are $(U^0, \ldots, U^n)$, corresponding to the coordinates on $\mathcal{X}$.

Configuration bundle. Above the base space we construct the ***configuration bundle*** $Y = \mathcal{X} \times \mathcal{S}$, which is the product of the base space $\mathcal{X}$ with the ambient space $\mathcal{S}$. This is an example of a ***fiber bundle*** over $\mathcal{X}$; take $\pi_{\mathcal{X}Y} : Y \to \mathcal{X}$ to be the projection map, and let coordinates on Y be $(X^0, X^1, \ldots, X^n, x^1, \ldots, x^m)$. We will use lowercase roman letters a, b, c to range over $1, \ldots, m$, so coordinates on Y can be written either as (t, X^i, x^a) or as (X^μ, x^a).

A ***configuration*** of the system is specified by a map $\phi : \mathcal{U} \to Y$ covering a map $\phi_{\mathcal{X}} : \mathcal{U} \to \mathcal{X}$. That is, ϕ satisfies $\pi_{\mathcal{X}Y} \circ \phi = \phi_{\mathcal{X}}$, so that $\phi(U) = (\phi^\mu(U), \phi^a(U))$. The map ϕ is taken to be smooth and ϕ_X is assumed to be a diffeomorphism, so that it is smooth with a smooth inverse. The exact class of regularity will not be of importance at the moment, but of course such notions are crucial for analytical studies, including error estimates.

We will frequently be interested in the composition $\varphi = \phi \circ \phi_{\mathcal{X}}^{-1} : \phi_{\mathcal{X}}(\mathcal{U}) \subset \mathcal{X} \to Y$ which maps a time t and a material position X to the corresponding deformed position x. The fiber component of this map is thus exactly the deformation mapping, and we have the following commutative diagram



A deformation mapping is thus a ***section*** of the configuration bundle, defined over all space and time, meaning that $\pi \circ \varphi = \mathrm{id}$. This is shown graphically in Fig. 5.1, where the section is regarded as a surface in the fiber bundle over the base space.

Jet bundle. Given a configuration bundle Y over a base space $\mathcal{X}$, we next construct the ***jet bundle*** J^1Y over Y with fibers over x_X consisting of linear maps $\gamma : T_X\mathcal{X} \to T_xY$ such that $T\pi_{XY} \cdot \gamma = \mathrm{id}_{\mathcal{X}}$. This is the space of partial derivatives with respect to space and time (space-time velocities). Coordinates on J^1Y are denoted $(X^\mu, x^a, v^a{}_\mu) \equiv (t, X^i, x^a, v^a{}_t, v^a{}_i)$. When we are writing time and space coordinates separately, we will use (t, X, x, v_t, v_X) to indicate the time and space partial derivatives.

Given a section φ of Y, $T_X\varphi$ is an element of $(J^1Y)_X$, and we define the ***jet extension*** of φ to be $j^1\varphi : X \to (X, T_X\varphi)$. This is φ together with its partial derivatives and in coordinates it is written $j^1\varphi(X) = (X^\mu, \varphi^a(X), \varphi^a{}_{,\mu}(X))$, where we denote the partial derivatives by $\varphi^a{}_{,\mu}(X) = \frac{\partial \varphi^a}{\partial X^\mu}(X)$. We use (X, x, v) to refer to a general point in J^1Y, and $j^1\varphi(X) = (X, \varphi(X), \varphi_{,X}(X))$ to refer to a point which comes from the first jet of a section. A jet extension is thus an example of a section of the fiber bundle $J^1Y \to X$.

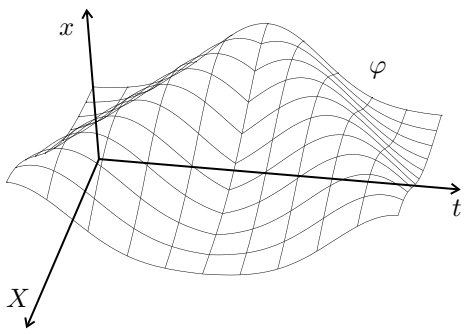


Figure 5.1: A graphical representation of a section φ of a bundle for elastodynamics. The horizontal axes represent space-time and together they form the base space $\mathcal{X} = \mathbb{R} \times \mathbb{R}^3$. The vertical axis represents the ambient space, so the entire bundle is $\mathcal{S} \times \mathcal{X}$. Taking a slice of φ with constant $X \in \mathbb{R}^3$ gives the trajectory of the particle with material coordinates X for all time. Alternatively, taking a slice of φ with constant $t \in \mathbb{R}$ gives the configuration of the entire body at a single instant of time.

In the terminology of solid mechanics, the time component of the first jet of a section is the material velocity and the space components form the deformation gradient. That is,

$$v_t = \dot{\varphi}(X) \qquad \text{and} \qquad v_X = F(X),$$

where $(X, x, v) = j^1\varphi(X)$.

Note that J^1Y is not the tangent bundle TY of Y. It is also not the tangent bundle TS, as this would only include one derivative (for example, with respect to time) of a configuration, whereas the each element of the jet bundle includes the derivatives with respect to all the base space coordinates (space and time).

Lagrangian. To define a particular system it is necessary to specify a ***Lagrangian*** $L : J^1Y \to \mathbb{R}$, which maps the first jet bundle to the real numbers. For continuum mechanics the Lagrangian has the form

$$L(t, X, x, v_t, v_X) = \left[\frac{1}{2}\rho(X)\|v_t\|^2\right] - \left[W(X, v_X) + \rho(X)V(X, x)\right], \tag{5.1}$$

where $\rho : \mathcal{B} \to \mathbb{R}$ is the (material) ***density***, $W : (X, v_X) \to \mathbb{R}$ is the ***stored energy function*** per unit volume and $V : Y \to \mathbb{R}$ is the ***external potential function*** per unit mass. Different forms of W determine the different types of continua, such as fluids and solids, while V specifies the environmental potentials such as gravity. The external potential V specifies body forces of potential type by $\mathbf{B} = -\nabla V$. The two terms in the Lagrangian (5.1) correspond to the kinetic and potential energy, respectively.

Unlike the standard Lagrangians or Hamiltonians used for continuum mechanics, the multisym-

plectic Lagrangian is purely local. This is an explicit formulation of the fact that classical continuum theories do not involve long-range dependencies in their constitutive or geometric foundations.

An intrinsic formulation of multisymplectic mechanics of continua (such as that in Marsden et al. [2001]) is based on the ***Lagrangian density***, which is a map from J^1Y to the space $\Lambda^{n+1}(\mathcal{X})$ of volume densities on $\mathcal{X}$. To form a Lagrangian density from our Lagrangian, simply take $L\mathbf{d}^{n+1}X$, where $\mathbf{d}^{n+1}X$ is the standard volume element on $\mathbb{R}^{n+1}$.

Dual jet bundle. We now briefly consider the Hamiltonian viewpoint of multisymplectic field theories. The approach taken here is non-intrinsic, and we are thus neglecting much of the geometry underlying such systems. The interested reader is referred to Marsden and Shkoller [1999] for an intrinsic formulation of multisymplectic Hamiltonian mechanics and to Marsden et al. [2001] for the special case of continuum mechanics.

For multisymplectic mechanics, the natural dual to the jet bundle is the ***affine dual*** $J^1Y^\star$, with coordinates $(X^\mu, x^a, {p_a}^\mu, p)$, representing the map ${v^a}_\mu \mapsto p + {p_a}^\mu {v_a}^\mu$. Here ${p_a}^\mu$ are the space-time momenta, and p is an additional scalar, which we will see is related to the energy. The need to consider the affine dual, rather than the linear dual as in classical mechanics, becomes apparent when one considers Noether's theorem for multisymplectic mechanics. Note that $J^1Y^\star$ is not the cotangent bundle T^*Y of Y.

Legendre transform. Given a Lagrangian L on a jet bundle $J^1Y \to \mathcal{X}$, we construct a map from the jet bundle to the dual jet bundle known as the ***Legendre transform*** $\mathbb{F}L : J^1Y \to J^{1*}Y$. It is defined by

$$\mathbb{F}L : (X^\mu, x^a, {v^a}_\mu) \mapsto (X^\mu, x^a, {p_a}^\mu, p), \tag{5.2}$$

where

$${p_a}^\mu = \frac{\partial L}{\partial {v^a}_\mu}(x, y, v) \qquad \text{and} \qquad p = {p^a}_\mu {v_a}^\mu - L(X, x, v).$$

Calculating the Legendre transform for the continuum mechanics Lagrangian (5.1) gives

$${p_a}^t = \rho(x){v^a}_t \tag{5.3a}$$

$${p_a}^i = -{P_a}^i(X) \tag{5.3b}$$

$$p = \left[\frac{1}{2}\rho(X)\|v_t\|^2\right] + [W(X, v_X) + \rho(X)V(X, x)] - F(X) : P(X). \tag{5.3c}$$

We see here that the time momenta are the classical momentum values, while the space momenta are the (negative of) the first Piola-Kirchhoff stress tensor.

If $\mathbb{F}L$ has maximal rank at some point in J^1Y, then the Lagrangian is said to be ***regular*** at that point. Note that this does not imply that $\mathbb{F}L$ is a local isomorphism, as the dimension of the

dual jet bundle is one more than that of the jet bundle, and so the Legendre transform can never be surjective.

The Legendre transform can also be used to define the ***energy*** function $E_L : J^1Y \to \mathbb{R}$ associated to a Lagrangian L by

$$\begin{aligned} E_L(X, x, v) &= p^a{}_t v_a{}^t - L(X, x, v) \\ &= \left[\frac{1}{2}\rho(X)\|v_t\|^2\right] + \left[W(X, v_X) + \rho(X)V(X, x)\right], \end{aligned}$$

where $(X, x, p) = \mathbb{F}L(X, x, v)$. This will be important later when we consider conservation laws for Lagrangian systems.

5.1.2 Variations and dynamics

Configuration space and variations. Take $\mathcal{C}(Y)$ to be the space of all configurations ϕ. We will frequently wish to consider variations of solutions, which are tangent vectors to a smooth curve of configurations. To define these, first consider the tangent bundle TY of Y, which has coordinates $(X, x, \delta X, \delta x)$.

Using this, we see that the tangent space to $\mathcal{C}(Y)$ at a configuration ϕ is denoted $T_\phi\mathcal{C}(Y)$ and consists of all maps $\delta\phi : \mathcal{U} \to TY$ of the form $\delta\phi(U) = (\phi^\mu, \phi^a, \delta\phi^\mu, \delta\phi^a)$. Such tangent vectors are called ***variations*** of the configuration ϕ. The components $\delta\phi^a$ are termed ***vertical variations***, while the $\delta\phi^\mu$ are called ***horizontal variations***. While the definition of a vertical variation is well defined, selecting a particular direction for horizontal variations requires additional structure on the configuration bundle. Here we have implicitly assumed this by working in a preferred set of coordinates. An intrinsic alternative can also be provided by taking horizontal variations to be those which are tangent to $j^1(\phi \circ \phi_{\mathcal{X}}^{-1})$ (see Marsden and Shkoller [1999]).

Euler-Lagrange equations. Given the configuration space $\mathcal{C}(Y)$ of all possible ϕ, it is necessary to determine which of these configurations will be adopted by the system. To do this, we introduce the ***action integral*** $S : \mathcal{C}(Y) \to \mathbb{R}$, defined as

$$S(\phi) = \int_{\phi_{\mathcal{X}}(\mathcal{U})} L\Big(j^1(\phi \circ \phi_{\mathcal{X}}^{-1})\Big)\, dV, \tag{5.4}$$

where dV is the volume element on $\mathcal{X}$.

Note that $S(\phi)$ only depends on ϕ through φ, so that for any diffeomorphism $\gamma : \mathcal{U} \to \mathcal{U}$, $S(\phi\circ\gamma) = S(\phi)$. We will see later that this implies that the Euler-Lagrange equations only determine φ uniquely, rather than the full ϕ.

Hamilton's principle now states that the physical configurations ϕ are those which are *critical*

points of the action function. More precisely, Hamilton's principle requires that

$$\mathbf{d}S(\phi)\cdot\delta\phi = 0 \tag{5.5}$$

for all variations $\delta\phi \in T_\phi\mathcal{C}(Y)$ which are zero on the boundary $\partial\mathcal{U}$ of $\mathcal{U}$. This is the classical ***weak form*** of the equation.

To derive the ***strong form***, we first rewrite the action as

$$S(\phi) = \int_{\mathcal{U}} L\left(\phi^\mu(U), \phi^a(U), \frac{\partial\phi^a}{\partial U}\cdot\left[\frac{\partial\phi_{\mathcal{X}}}{\partial U}\right]^{-1}\right)\det\left[\frac{\partial\phi_{\mathcal{X}}}{\partial U}\right]\,dU$$

and now we compute $\mathbf{d}S$ to obtain

$$\begin{aligned}
&\mathbf{d}S(\phi)\cdot\delta\phi \\
&\quad= \int_{\mathcal{U}}\Bigg[\Bigg(\frac{\partial L}{\partial X^\mu}\delta\phi^\mu + \frac{\partial L}{\partial x^a}\delta\phi^a \\
&\qquad\qquad + \frac{\partial L}{\partial v^a{}_\mu}\left[\frac{\partial\delta\phi^a}{\partial U^\nu}J^\nu{}_\mu - \frac{\partial\phi^a}{\partial U^\nu}J^\nu{}_\rho\frac{\partial\delta\phi^\rho}{\partial U^\gamma}J^\gamma{}_\mu\right]\Bigg)\det\left[\frac{\partial\phi_{\mathcal{X}}}{\partial U}\right] \\
&\qquad\qquad + L\det\left[\frac{\partial\phi_{\mathcal{X}}}{\partial U}\right]J^\nu{}_\mu\frac{\partial\delta\phi^\mu}{\partial U^\nu}\Bigg]\,dU \\
&\quad= \int_{\phi_{\mathcal{X}}(\mathcal{U})}\Bigg(\left[\frac{\partial L}{\partial x^a} - \frac{d}{dX^\mu}\left(\frac{\partial L}{\partial v^a{}_\mu}\right)\right]\delta\phi^a \\
&\qquad\qquad + \left[\frac{\partial L}{\partial X^\nu} + \frac{d}{dX^\mu}\left(\frac{\partial L}{\partial v^a{}_\mu}\frac{\partial\varphi^a}{\partial X^\nu}\right) - \frac{dL}{dX^\nu}\right]\delta\phi^\nu\Bigg)\,dX \\
&\qquad + \int_{\partial\phi_{\mathcal{X}}(\mathcal{U})}\left(\frac{\partial L}{\partial v^a{}_\mu}\delta\phi^a N_\mu - \left[\frac{\partial L}{\partial v^a{}_\mu}\frac{\partial\varphi^a}{\partial X^\nu} - L\delta^\mu_\nu\right]\delta\phi^\nu N_\mu\right)\,dA,
\end{aligned} \tag{5.6}$$

where

$$J^\nu{}_\rho = \left[\left(\frac{\partial\phi_{\mathcal{X}}}{\partial U}\right)^{-1}\right]^\nu_\rho$$

and we have written φ instead of ϕ when taking derivatives with respect to X.

Restricting to variations which are zero on the boundary of $\mathcal{U}$ eliminates the boundary term from the above expression, and then requiring that it is zero for all such variations implies that both components of the integrand in the above expression must be zero. The first of these is the ***Euler-Lagrange equations***, which are

$$\frac{\partial L}{\partial x^a}(j^1\varphi(X)) - \frac{\partial}{\partial X^\mu}\left[\frac{\partial L}{\partial v^a{}_\mu}(j^1\varphi(X))\right] = 0 \quad \text{for all } X\in\mathcal{X}. \tag{5.7}$$

This is a PDE with dependent variables φ^a and independent variables X^μ. Indeed, as we will see

below, the second term in (5.6) is zero whenever the Euler-Lagrange equations are zero, which is the reason that the Euler-Lagrange equations are sufficient to describe the motion of the system.

For the continuum mechanics Lagrangian (5.1), the Euler-Lagrange equations are

$$\rho(X)\varphi^a{}_{,tt} = \frac{\partial}{\partial X^\mu}\left[\frac{\partial W}{\partial v^a{}_\mu}(X, \varphi_{,X}(X))\right] - \rho(X)\frac{\partial V}{\partial x^a}(X, \varphi(X)). \tag{5.8}$$

Equations of motion. Substituting the definitions for the material velocity and first Piola-Kirchhoff stress tensor into the Euler-Lagrange equations (5.8) gives the familiar equation

$$\rho\varphi_{,tt} = \mathrm{DIV}P - \rho\nabla_X V. \tag{5.9}$$

The term $-\nabla V$ is simply the external ***body forces***, which is often expressed as $\mathbf{B}(X,t)$. If there are non-potential forces present, these are added to the right hand side of (5.9).

Boundary conditions. For first-order multisymplectic theories we consider only zeroth- or first-order boundary conditions. That is, we allow boundary conditions of the form

$$\varphi(\phi_{\mathcal{X}}(U)) = \varphi_0(U) \quad \text{for } U \in \partial_0\mathcal{U} \tag{5.10a}$$

$$\frac{\partial L}{\partial v^a{}_\mu}N_\mu(\phi_{\mathcal{X}}(U)) = \tau_a(U) \quad \text{for } U \in \partial_1\mathcal{U}, \tag{5.10b}$$

where $\partial_0\mathcal{U}$ and $\partial_1\mathcal{U}$ are subsets of the boundary $\partial\mathcal{U}$, φ_0 is a given section, τ is a given one-form and $N_\mu(X)$ is the normal one-form to the boundary $\phi_{\mathcal{X}}(\partial\mathcal{U})$. We say that (5.10a) is a ***zeroth-order*** boundary condition, whereas (5.10b) is a ***first-order*** boundary condition. For the moment, we do not require that $\partial_0\mathcal{U}$ and $\partial_1\mathcal{U}$ be disjoint, nor do we require that their union cover $\partial\mathcal{U}$, although such conditions on the partitions of $\partial\mathcal{U}$ become important for well-posedness.

As in standard Lagrangian theories, we can either impose the boundary conditions (5.10a) and (5.10b) directly, or we can modify Hamilton's principle (5.5) and then derive the boundary conditions from the variational principle. To do this, we say that ϕ is a solution *satisfying the boundary conditions* if

$$\mathbf{d}S(\phi)\cdot\delta\phi = \int_{\phi_{\mathcal{X}}(\partial_1\mathcal{U})} \tau_a\delta\phi^a \; dA \tag{5.11}$$

for all variations $\delta\phi$ which are zero on the set $\partial\mathcal{U}\setminus\partial_1\mathcal{U}$, and where we only consider sections ϕ which satisfy (5.10a).

Note that this is only one possible approach. It is also common to include in the potential energy a term whose derivative gives the traction boundary conditions. That approach is simpler, but the additional potential term is not intrinsic, whereas the expression (5.11) is intrinsically well defined.

Computing the left-hand side of (5.11) and using integration by parts gives (5.6). The boundary

term can be taken only over $\partial_1\mathcal{U}$ as $\delta\phi$ is zero elsewhere on $\partial\mathcal{U}$, and this matches with the right hand side of (5.11) to imply the traction boundary condition (5.10b). The displacement boundary condition (5.10a) is satisfied by assumption. As the set of variations $\delta\phi$ which are zero on all of $\partial\mathcal{U}$ is a subset of those we are using here, we also recover the Euler-Lagrange equations (5.7) from the variational principle with boundary terms (5.11).

For continuum mechanics we are particularly interested in the case of an initial boundary value problem. Recall that our parameter space is $\mathcal{U} = [0,T] \times \mathcal{B}$ and that the boundary is therefore $\partial\mathcal{U} = (\{0\} \times \mathcal{B}) \cup (\{T\} \times \mathcal{B}) \cup ([0,T] \times \partial\mathcal{B})$. An initial boundary value problem specifies that

$$\varphi^a(\phi_\mathcal{X}(0,U_X)) = (\varphi_0)^a(0,U_X) \qquad \text{for all } U_X \in \mathcal{B} \tag{5.12a}$$

$$\varphi^a{}_{,t}(\phi_\mathcal{X}(0,U_X)) = (\varphi_0)^a{}_{,t}(0,U_X) \qquad \text{for all } U_X \in \mathcal{B} \tag{5.12b}$$

$$\varphi^a(\phi_\mathcal{X}(U_t,U_X)) = (\varphi_0)^a(U_t,U_X) \qquad \text{for all } U_t \in [0,T],\ U_X \in \partial_d\mathcal{B} \tag{5.12c}$$

$$\frac{\partial L}{\partial v^a{}_i} N_i(\phi_\mathcal{X}(U_t,U_X)) = -T_a(U_t,U_X) \qquad \text{for all } U_t \in [0,T],\ U_X \in \partial_\tau\mathcal{B}, \tag{5.12d}$$

where φ_0 and T_a are given functions on $\mathcal{U}$ and $\partial_d\mathcal{B}$ and $\partial_\tau\mathcal{B}$ are disjoint subsets of $\partial\mathcal{B}$ whose union covers $\partial\mathcal{B}$. The first two conditions (5.12a) and (5.12b) are the ***initial conditions***, while (5.12c) and (5.12d) are the ***boundary conditions***.

In terms of the conditions (5.10), we identify the zeroth- and first-order boundary conditions as defined on

$$\partial_0\mathcal{U} = (\{0\} \times \mathcal{B}) \cup ([0,T] \times \partial_d\mathcal{B})$$
$$\partial_1\mathcal{U} = (\{0\} \times \mathcal{B}) \cup ([0,T] \times \partial_\tau\mathcal{B}).$$

Note that these sets are neither disjoint nor covering.

5.1.3 Horizontal variations

Requiring stationarity with respect to horizontal variations implies that the second term in (5.6) must be zero, which gives

$$\frac{\partial L}{\partial X^\nu} + \frac{d}{dX^\mu}\left(\frac{\partial L}{\partial v^a{}_\mu}\frac{\partial \varphi^a}{\partial X^\nu}\right) - \frac{dL}{dX^\nu} = 0. \tag{5.13}$$

While it might initially seem that $\mathbf{d}S(\phi)\cdot\delta\phi = 0$ for all $\delta\phi$ zero on $\partial\mathcal{U}$ would require that both the Euler-Lagrange equations (5.7) and the equation (5.13) are satisfied, in fact it is sufficient to require that only the Euler-Lagrange equations are satisfied. The reason for this is that equation (5.13) is

implied by the Euler-Lagrange equations, as can be seen by calculating:

$$\begin{aligned}\frac{\partial L}{\partial X^\nu} + \frac{d}{dX^\mu}\left(\frac{\partial L}{\partial {v^a}_\mu}\frac{\partial \varphi^a}{\partial X^\nu}\right) - \frac{dL}{dX^\nu} & \\ &= \frac{\partial L}{\partial X^\nu} + \frac{d}{dX^\mu}\left(\frac{\partial L}{\partial {v^a}_\mu}\right)\frac{\partial \varphi^a}{\partial X^\nu} + \frac{\partial L}{\partial {v^a}_\mu}\frac{d}{dX^\mu}\left(\frac{\partial \varphi^a}{\partial X^\nu}\right) \\ &\qquad - \left[\frac{\partial L}{\partial X^\nu} + \frac{\partial L}{\partial \varphi^a}\frac{\partial \phi^a}{\partial X^\nu} + \frac{\partial L}{\partial {v^a}_\nu}\left(\frac{\partial \varphi^a}{\partial X^\mu}\right)\right] \\ &= -\left[\frac{\partial L}{\partial \phi^a} - \frac{d}{dX^\mu}\left(\frac{\partial L}{\partial {v^a}_\mu}\right)\right]\frac{\partial \varphi^a}{\partial X^\nu}\end{aligned}$$

and thus we see that whenever the Euler-Lagrange equations are satisfied, so too is equation (5.13). This can also be understood as a reflection of the symmetry of the action under the transformation $\phi \mapsto \phi \circ \gamma$. Equation (5.13) is exactly Noether's theorem for this action. By now considering the space and time components of (5.13) separately, we will next see that this is in fact a restatement of very well-known facts about solutions of the equations of motion.

Energy conservation. Considering the special case of the base space X being space-time, the time component of the equation (5.13) is

$$\frac{\partial L}{\partial t} + \frac{d}{dt}\left(\frac{\partial L}{\partial {v^a}_t}\dot{\varphi}^a - L\right) + \frac{d}{dX^i}\left(\frac{\partial L}{\partial {v^a}_i}\dot{\varphi}^a\right) = 0,$$

which is the energy evolution equation. Assuming that $\phi_{\mathcal{X}} = \mathrm{id}$, in the special case that L does not depend explicitly on t we can integrate over the material body to obtain

$$\begin{aligned}\frac{d}{dt}\int_{\mathcal{B}}\left(\frac{\partial L}{\partial {v^a}_t}\dot{\varphi}^a - L\right)\,dV &= -\int_{\mathcal{B}}\frac{d}{dX^i}\left(\frac{\partial L}{\partial {v^a}_i}\dot{\varphi}^a\right)\,dV \\ &= -\int_{\partial\mathcal{B}}\frac{\partial L}{\partial {v^a}_i}\dot{\varphi}^a N_i\,dA \\ &= -\int_{\partial_\tau\mathcal{B}}\tau_a\dot{\varphi}^a\,dA.\end{aligned}$$

In the particular case of traction-free boundary conditions, when $\tau = 0$ on $\partial\mathcal{B}$, then this reduces to

$$\frac{d}{dt}\int_{\mathcal{B}}\left(\frac{\partial L}{\partial {v^a}_t}\dot{\varphi}^a - L\right)\,dV = 0, \tag{5.14}$$

which is the statement of global energy conservation. As we will see below, this calculation can also be recast in the form of Noether's theorem for horizontal symmetry actions.

Configurational forces. Having considered the time component of equation (5.13) above, we now consider the full expression

$$\frac{\partial L}{\partial X^\nu} + \frac{d}{dX^\mu}\left(\frac{\partial L}{\partial v^a{}_\mu}\frac{\partial \varphi^a}{\partial X^\nu} - L\delta^\mu_\nu\right) = 0. \tag{5.15}$$

In this equation we can recognize the Eshelby Energy-Momentum tensor $\mathbf{C}$ (see, e.g., Gurtin [2000])

$$C^\mu_\nu = \frac{\partial L}{\partial v^a{}_\mu}\frac{\partial \varphi^a}{\partial X^\nu} - L\delta^\mu_\nu$$

and equation (5.15) expresses the balance of the configurational forces. Surface independent integrals, such as the static and dynamic J-integrals, are obtained from it. These appear whenever $\frac{\partial L}{\partial X^i} = 0$ by integrating over an arbitrary volume and using Stokes' theorem to transform it into a boundary integral. In the two-dimensional case, these integrals are path integrals.

5.2 Conservation laws

One of the primary advantages of multisymplectic theories is the clear understanding which can be gained from the conservation laws satisfied by the system. As we shall see, all conservation laws considered here can be expressed in either a local divergence form or in a global form.

Space of solutions. To understand both local and global statements of conservation laws it is necessary to take variations and divergences along solutions.

Recall that we are using $\mathcal{C}(Y)$ to denote the space of all configurations $\phi : \mathcal{U} \to Y$. The ***space of solutions*** $\mathcal{C}_L(Y) \subset \mathcal{C}(Y)$ is the subset which is composed of those ϕ which satisfy the Euler-Lagrange equations everywhere, for any boundary conditions. That is, $\mathcal{C}_L(Y)$ is the set of solutions for all possible choices of boundary conditions. As we have already remarked, the fact that the action (5.4) only depends on ϕ via φ means that solutions $\phi \in \mathcal{C}_L(Y)$ are only unique up to reparameterization $\phi \circ \gamma$ for diffeomorphisms $\gamma : \mathcal{U} \to \mathcal{U}$.

The tangent bundle of the space of solutions is denoted $T\mathcal{C}_L(Y)$, and a variation $V \in T_\phi\mathcal{C}_L(Y)$ is thus the derivative of a curve of solutions, typically having different boundary data. Such V are known as ***first variations*** of ϕ. In fact $\mathcal{C}_L(Y)$ may not be a smooth manifold (see, for example, Fischer, Marsden, and Moncrief [1980] and Arms, Marsden, and Moncrief [1982]) and so a more general definition of first variations should be used. Here we will assume smoothness, and we refer the reader to Marsden et al. [1998] for the details of the general case.

Local actions. In what follows it will frequently be convenient to consider the action integral taken over a subset $\mathcal{U}'$ of $\mathcal{U}$. We will denote this by $S'(\phi)$, so that

$$S'(\phi) = \int_{\phi_X(\mathcal{U}')} L\Big(j^1(\phi \circ \phi_{\mathcal{X}}^{-1})\Big)\, dV.$$

5.2.1 Multisymplectic forms

In this section we introduce the multisymplectic structures which give multisymplectic mechanics its name. This can be done in two ways, either on the Lagrangian side from the variational principle, or on the Hamiltonian side by direct construction. We will consider only the Lagrangian side of the picture, and we refer to Marsden and Shkoller [1999] for a comparison of the Lagrangian and Hamiltonian constructions. For simplicity, the material here is a non-intrinsic version of the theory developed in Gotay et al. [1997] and Marsden et al. [1998].

Given a variation $V : \mathcal{U} \to TY$ of a configuration ϕ we denote by $j^1V : \mathcal{U} \to T(J^1Y)$ its jet prolongation. If ϕ^ϵ is a smooth curve from $\mathbb{R}$ to $\mathcal{C}(Y)$ such that

$$V = \left.\frac{\partial \phi^\epsilon}{\partial \epsilon}\right|_{\epsilon=0} \text{ and } \phi^0 = \phi,$$

then the jet prolongation of V is defined by

$$j^1V = \left.\frac{\partial j^1\phi^\epsilon}{\partial \epsilon}\right|_{\epsilon=0}.$$

In coordinates this is given by

$$j^1V(U) = \left(\phi^\mu, \phi^a, V^\mu, V^a, \frac{\partial V^a}{\partial X^\mu} - \frac{\partial V^\nu}{\partial X^\mu} v^a{}_\nu\right).$$

Free action variations. For a variational derivation of the multisymplectic structure, we return to the variational principle and consider the expression $\mathbf{d}S(\phi) \cdot \delta\phi$ for arbitrary $\delta\phi$. That is, we do not require that $\delta\phi$ vanishes on the boundary $\partial\mathcal{U}$, so we have the full expression (5.6) for action variations.

Multisymplectic $n+1$-form. We now restrict ourselves to configurations $\phi \in \mathcal{C}_L(Y)$ which are solutions of the Euler-Lagrange equations, and thus also satisfy the horizontal equation (5.13), and we consider variations V which lie in the tangent space $T\mathcal{C}_L(Y)$ of the space of solutions. This means that the first integral in the above expression is identically zero, and working with an arbitrary $\mathcal{U}' \subset \mathcal{U}$ we can write

$$\mathbf{d}S'(\phi) \cdot V = \int_{\partial\phi_{\mathcal{X}}(\mathcal{U}')} (j^1(\phi \circ \phi_{\mathcal{X}}^{-1}))^* \Big(\mathbf{i}_{j^1V}\Theta_L\Big), \tag{5.16}$$

where the ***Lagrangian*** $n+1$***-form*** Θ_L on J^1Y is defined by

$$\Theta_L = \frac{\partial L}{\partial v^a{}_\mu}\mathbf{d}x^a \wedge \mathbf{d}^n X_\mu - \left(\frac{\partial L}{\partial v^a{}_\mu}v^a{}_\mu - L\right)\mathbf{d}^{n+1}X.$$

Here we use the notation from Marsden et al. [1998], in which $\mathbf{d}^{n+1}X$ is the volume form on $\mathcal{X}$ and $\mathbf{d}^n X_\mu = \mathbf{i}_{\partial/\partial\mu}\mathbf{d}^{n+1}X$ are a set of n-forms. This is related to the previous expression for $\mathbf{d}S$ by the fact that $\mathbf{i}_V\mathbf{d}^{n+1}X = V \cdot N\mathbf{d}A$ on a surface with area element A induced from the volume form.

The fact that Θ_L has degree higher than one is one reason for the 'multi' in the term 'multisymplectic'. Another interpretation of this term, used in Bridges [1997] and Bridges and Reich [2001a], arises from defining the vector valued one-forms

$$\Theta_L^\mu = \frac{\partial L}{\partial v^a{}_\mu}\mathbf{d}y^a$$

for each $\mu = 1, \ldots, n$. For vertical first variations V, we can then write the derivative of the action as

$$\mathbf{d}S'(\phi)\cdot V = \int_{\phi_{\mathcal{X}}(\partial\mathcal{U}')} \Theta_L^\mu \cdot j^1V \; dA,$$

where we are somewhat loose about the precise meaning of this expression. The fact that there are $n+1$ different one-forms Θ_L^μ gives a second meaning to the prefix 'multi'. Note, however, that this decomposition into $n+1$ one-forms depends on the choice of coordinates and so is not intrinsic, whereas Θ_L is.

Multisymplectic $n+2$-form. Having derived the Lagrangian $n+1$-form as the boundary terms in the variations of the actions, we can now take the exterior derivative to obtain the ***multisymplectic Lagrangian*** $n+2$***-form***

$$\Omega_L = -\mathbf{d}\Theta_L.$$

We will shortly see why this is an important object. This can be written as

$$\Omega_L = \mathbf{d}x^a \wedge \mathbf{d}\left(\frac{\partial L}{\partial v^a{}_\mu}\right) \wedge \mathbf{d}^n X_\mu + \mathbf{d}\left(\frac{\partial L}{\partial v^a{}_\mu}v^a{}_\mu - L\right) \wedge \mathbf{d}^{n+1}X,$$

where $\mathbf{d}^n X_\mu$ and $\mathbf{d}^{n+1}X$ are as defined above. Fully expanded in coordinates, this becomes

$$\begin{aligned}\Omega_L = \frac{\partial^2 L}{\partial X^\mu \partial v^a{}_\mu}\mathbf{d}x^a \wedge \mathbf{d}^{n+1}X + \frac{\partial^2 L}{\partial x^b \partial v^a{}_\mu}\mathbf{d}x^a \wedge \mathbf{d}x^b \wedge \mathbf{d}^n X_\mu \\ + \frac{\partial^2 L}{\partial v^b{}_\nu \partial v^a{}_\mu}\mathbf{d}x^a \wedge \mathbf{d}v^b{}_\nu \wedge \mathbf{d}^n X_\mu + \frac{\partial^2 L}{\partial x^b \partial v^a{}_\mu}v^a{}_\mu \mathbf{d}x^b \wedge \mathbf{d}^{n+1}X \\ + \frac{\partial^2 L}{\partial v^b{}_\nu \partial v^a{}_\mu}v^a{}_\mu \mathbf{d}v^b{}_\nu \wedge \mathbf{d}^{n+1}X - \frac{\partial L}{\partial x^a}\mathbf{d}x^a \wedge \mathbf{d}^{n+1}X.\end{aligned}$$

5.2.2 Multisymplectic form formula

Now that we have defined the multisymplectic forms, we will derive the conservation properties associated with them.

Recall that the exterior derivative satisfies $\mathbf{d}^2 = 0$. For Euclidean (flat) spaces, this can be written $\mathbf{d}^2 S(V,W) = D(DS \cdot W) \cdot V - D(DS \cdot V) \cdot W$, where D denotes the Fréchet derivative. This expression is zero as the partial derivatives commute, although it is also true in more general non-flat settings as well.

We can now use this fact to take a second exterior derivative of the identity (5.16) restricted to the space of solutions $\mathcal{C}_L(Y)$ and conclude that it must be zero. The intrinsic calculation of this (see Marsden et al. [1998]) gives the ***multisymplectic form formula***

$$\mathbf{d}^2 S'(\phi)(V,W) = \int_{\partial\phi_{\mathcal{X}}(\mathcal{U}')} (j^1(\phi \circ \phi_{\mathcal{X}}^{-1}))^* \Big(\mathbf{i}_{j^1 W} \mathbf{i}_{j^1 V} \Omega_L \Big) = 0 \tag{5.17}$$

for all first variations V and W of a solution ϕ. This is the ***global*** form of the multisymplectic conservation law.

Applying Stokes' theorem and using the fact that $\mathcal{U}'$ is arbitrary implies that the above statement is equivalent to the ***local multisymplectic form formula***

$$\mathbf{d} \left[(j^1(\phi \circ \phi_{\mathcal{X}}^{-1}))^* \Big(\mathbf{i}_{j^1 W} \mathbf{i}_{j^1 V} \Omega_L \Big) \right] = 0, \tag{5.18}$$

where V and W are again first variations of a solution ϕ. This statement holds at every point in $\mathcal{U}$ or, equivalently, in $\mathcal{X}$.

As mentioned earlier, the above results cannot in fact be obtained simply by taking exterior derivatives of (5.16), as the space of solutions may not be a smooth manifold. This necessitates the use of a more general definition of a first variation, and somewhat complicates the proof of the multisymplectic form formula. We refer to Marsden et al. [1998] for the details.

Note that here we do not appear to have explicitly considered initial or boundary conditions. This is because the variations V and W implicitly contain variations in the initial or boundary conditions, as these conditions act as a parameterization of the space of solutions $\mathcal{C}_L(Y)$ by distinguishing nearby solutions from each other (away from bifurcation points), up to reparameterization by diffeomorphisms $\gamma : \mathcal{U} \to \mathcal{U}$.

In the general case the coordinate expressions for the multisymplectic form formula are very complicated. If we restrict attention to only vertical variations, however, then we can write (5.17) explicitly, as we will now see.

5.2.3 Spatial multisymplectic form formula and reciprocity

We now turn to an explicit interpretation of the global multisymplectic form formula in the case of static continuum mechanics. As we shall see, in this particular case it is simply a restatement of the well-known Betti reciprocity theorem, when the variations are restricted to being purely vertical.

Linearized equations. Assume that $\phi_{\mathcal{X}} = \mathrm{id}$. Recall that we say that ϕ is a solution of the Euler-Lagrange equations with displacement and traction boundary conditions (5.12c) and (5.12d) if it satisfies

$$DS(\phi) \cdot V = \int_{\partial_\tau \mathcal{B}} \tau \cdot V \; dA$$

for all variations V which are zero on the displacement boundary $\partial_d \mathcal{B}$. We now define W to be a ***solution of the linearized problem*** at ϕ if

$$D(DS(\phi) \cdot V) \cdot W = 0$$

for all V vanishing on the displacement boundary. More generally, we say that W is a solution of the linearized problem with ***incremental body force*** $B(W)$ and ***incremental traction*** $\tau(W)$ if

$$D(DS(\phi) \cdot V) \cdot W = \int_{\mathcal{B}} B(W) \cdot V \; dV + \int_{\partial_\tau \mathcal{B}} \tau(W) \cdot V \; dA$$

for all V vanishing on $\partial_d \mathcal{B}$.

We now use the fact that for *any* two variations V and W, not necessarily vanishing anywhere, the multisymplectic form formula is simply the statement that $D(DS(\phi)\cdot V)\cdot W = D(DS(\phi)\cdot W)\cdot V$. This implies that

$$\int_{\mathcal{B}} B(W) \cdot V \; dV + \int_{\partial_\tau \mathcal{B}} \tau(W) \cdot V \; dA = \int_{\mathcal{B}} B(V) \cdot W \; dV + \int_{\partial_\tau \mathcal{B}} \tau(V) \cdot W \; dA,$$

which is exactly the statement of Betti reciprocity (see, for example, Marsden and Hughes [1994] or Truesdell and Noll [1965]).

In words, this means that if $B(W)$ and $\tau(W)$ are applied forces which produce the linearized response W, and $B(V)$ and $\tau(V)$ similarly produce V, then measuring the response V in the direction of the forces $B(W), \tau(W)$ gives the same answer as measuring the response W in the direction $B(V), \tau(V)$.

In classical mechanics it is also common to write a dynamic reciprocity theorem which holds at a given instant of time (see, for instance, Marsden and Hughes [1994]). This is done by including the linear momentum in the body force terms in the above system. This is not the same as a fully space-time reciprocity theorem, which can be derived exactly as above by simply considering a dynamic

problem and taking the action over the full space-time base space $[0,T]\times\mathcal{B}$. By taking space-time slices of the form $[t,t+(\Delta t)]\times\mathcal{B}$ and letting Δt go to zero, the fully space-time reciprocity theorem then can be used to derive the standard dynamic reciprocity theorem.

In general, reciprocity occurs in any system arising from a potential function. For an elegant general theory based on Lagrangian submanifolds see Marsden and Hughes [1994].

5.2.4 Temporal multisymplectic form formula and symplecticity

As we have seen above, reducing the multisymplectic form formula to only apply in space recovers the standard reciprocity theorem of elastostatics. We will now show how to recover the standard symplecticity relation of Hamiltonian or Lagrangian mechanics in time.

Assume that $\phi_{\mathcal{X}} = \mathrm{id}$. Recall that a Hamiltonian system on the cotangent bundle T^*Q of a configuration manifold Q with canonical symplectic structure $\mathbf{d}q^i \wedge \mathbf{d}p_i$ will have a flow map $F_H^t : T^*Q \to T^*Q$ which preserves this symplectic structure on T^*Q. The Lagrangian equivalent of this statement is that the Lagrangian flow map $F_L^t : TQ \to TQ$ on the tangent bundle TQ preserves the Lagrangian two-form $\mathbf{d}q^i \wedge \mathbf{d}\left(\frac{\partial L}{\partial \dot{q}^i}\right)$.

To see how this is a consequence of the multisymplectic form formula, we first define the ***instantaneous space of solutions*** to be $\mathcal{C}_{\mathcal{B}}(\mathcal{S}) = \{\varphi : \mathcal{B} \to \mathcal{S}\}$, which is the space of configurations at a given instant of time. The flow map of the system can now be written

$$F_L^t :\ T\mathcal{C}_{\mathcal{B}}(\mathcal{S}) \to T\mathcal{C}_{\mathcal{B}}(\mathcal{S}), \qquad (\varphi_0, \dot{\varphi}_0) \mapsto (\varphi_t, \dot{\varphi}_t),$$

where $\varphi_t(X)$ satisfies the Euler-Lagrange equations for some given boundary conditions (5.12c,5.12d) with $\tau = 0$ and the initial conditions $(\varphi_0, \dot{\varphi}_0)$.

If we now take the boundary conditions and the Lagrangian to be constant in time, and consider a variation $(\delta\varphi_0, \delta\dot{\varphi}_0)$ in the initial condition, then defining

$$V_t = T\pi_{\mathcal{B}} \cdot TF_L^t \cdot (\delta\varphi_0, \delta\dot{\varphi}_0),$$

where $\pi_{\mathcal{B}} : (\varphi, \dot{\varphi}) \mapsto \varphi$, we see that V is exactly a particular vertical first variation, in the sense of the previous sections. Note also that $\mathbf{d}S(\varphi)\cdot V$ will only consist of boundary integrals at the initial and final times, as V is a variation which preserves the boundary conditions and thus is zero on the displacement boundary of the reference configuration, while τ is zero on the traction boundary.

Constructing two such vertical first variations V and W and applying the multisymplectic form formula, we obtain

$$\int_{\mathcal{B}} \Omega^t_{L(0,X)}(j^1V(0,X), j^1W(0,X))\ dV - \int_{\mathcal{B}} \Omega^t_{L(T,X)}(j^1V(T,X), j^1W(T,X))\ dV = 0.$$

Recall, however, that $\Omega_L^t = \mathbf{d}q^a \wedge \mathbf{d}p_a{}^t$, and so we can rewrite the above expression as

$$\int_{\mathcal{B}} \Omega^t_{L(0,X)}\Big((V_0,\dot{V}_0),(W_0,\dot{W}_0)\Big)\ dV = \int_{\mathcal{B}} \Omega^t_{L(T,X)}\Big(TF_L^T\cdot(V_0,\dot{V}_0), TF_L^T\cdot(W_0,\dot{W}_0)\Big)\ dV,$$

where we have used the definition of the variations V and W as being induced from initial variations $(V_0,\dot{V}_0)$ and $(W_0,\dot{W}_0)$, respectively.

The left-hand side of the above expression is simply the field-theoretic Lagrangian two form on $T\mathcal{C}_{\mathcal{B}}(\mathcal{S})$, which is

$$\Omega_L^{FT} = \mathbf{d}\left(\int_{\mathcal{B}} \frac{\partial L}{\partial \dot{\varphi}^a}\ dV\right)\wedge \mathbf{d}\varphi^a,$$

whereas the right hand side is the pullback of this under the flow map. That is, we have derived the statement

$$\Omega_L^{FT} = (F_L^T)^*\Omega_L^{FT} \tag{5.19}$$

of time-symplecticity of the flow.

5.2.5 Noether's theorem

An important source of conservation laws in continuum mechanics is when there are symmetries in the system. Noether's theorem is the statement which relates a symmetry to the corresponding conserved quantity, and we will now show how this can be formulated within the context of variational multisymplectic mechanics, as in Gotay et al. [1997] and Marsden et al. [1998].

Consider a Lie group G with Lie algebra $\mathfrak{g}$ and identity e which acts on the left on Y according to $\Phi : G\times Y\to Y$ by diffeomorphisms $g : Y\to Y$ covering the action $\Phi^{\mathcal{X}} : G\times\mathcal{X}\to\mathcal{X}$ by diffeomorphisms $g_{\mathcal{X}} : \mathcal{X}\to\mathcal{X}$. That is, each element of G can be written as $g(X,x) = (g_{\mathcal{X}}(X), g_Y(X,x))$. The prolongation of the group action is $\Phi^{J^1Y} : G\times J^1Y\to J^1Y$ given by

$$g\cdot\gamma = Tg_Y\circ\gamma\circ Tg_{\mathcal{X}}^{-1},$$

which in coordinates is

$$g\cdot(X^\mu,x^a,v^a{}_\mu) = \left(g^\mu_{\mathcal{X}}(X), g^a_Y(X,x), \left[\frac{\partial g^a_Y}{\partial X^\nu} + \frac{\partial g^a_Y}{\partial x^b}v^b{}_\nu\right]\frac{\partial (g_{\mathcal{X}}^{-1})^\nu}{\partial X^\mu}\right).$$

This definition is chosen so that $j^1(g\circ\varphi\circ g_{\mathcal{X}}^{-1}) = g\circ j^1\varphi\circ g_{\mathcal{X}}^{-1}$. Given a group action and its prolongation, we next define the ***infinitesimal generators*** associated with a Lie algebra element $\xi\in\mathfrak{g}$ to be $\xi_{\mathcal{X}} : \mathcal{X}\to T\mathcal{X}$, $\xi_Y : Y\to TY$ and $\xi_{J^1Y} : J^1Y\to T(J^1Y)$, where

$$\xi_{\mathcal{X}}(X) = \left.\frac{d}{dg}\right|_{g=e}\left(\Phi_g^{\mathcal{X}}(X)\right)\cdot\xi$$

$$\xi_Y(y) = \left.\frac{d}{dg}\right|_{g=e} \left(\Phi_g(y)\right) \cdot \xi$$
$$\xi_{J^1Y}(\gamma) = \left.\frac{d}{dg}\right|_{g=e} \left(\Phi_g^{J^1Y}(\gamma)\right) \cdot \xi.$$

Computing the coordinate expressions for the infinitesimal generators gives

$$\xi_{\mathcal{X}}(X) = \left(X^\mu, \xi^\mu = \frac{\partial(\Phi_g^{\mathcal{X}})^\mu}{\partial g^m}\xi^m\right)$$
$$\xi_Y(X,x) = \left(X^\mu, x^a, \xi^\mu, \xi^a = \frac{\partial(\Phi_g)^a}{\partial g^m}(e)\xi^m\right)$$
$$\xi_{J^1Y}(X,x,v) = \left(X^\mu, x^a, {v^a}_\mu, \xi^\mu, \xi^a, {\xi^a}_\mu = \frac{\partial \xi_Y^a}{\partial x^b}{v^b}_\mu + \frac{\partial \xi_Y^a}{\partial X^\mu} - \frac{\partial \xi_Y^\nu}{\partial X^\mu}{v^a}_\nu\right).$$

If the symmetry action is purely vertical or purely horizontal, then the above coordinate expressions simplify somewhat.

We now define the ***Lagrangian momentum map*** (sometimes called the ***multimomentum map***) $J_L : J^1Y \to \mathfrak{g}^* \otimes \Lambda^n(J^1Y)$ to be

$$J_L(\xi) = \mathbf{i}_{\xi_{J^1Y}}\Theta_L, \tag{5.20}$$

where $\mathfrak{g}^*$ is the dual of the Lie algebra $\mathfrak{g}$ of G and $\Lambda^n(J^1Y)$ is the space of n-forms on J^1Y. In coordinates, this reads

$$J_L(\xi) = \left(\frac{\partial L}{\partial {v^a}_\mu}\xi_Y^a - \left[\frac{\partial L}{\partial {v^a}_\nu}{v^a}_\nu - L\right]\xi_Y^\mu\right)\mathbf{d}^nX_\mu - \frac{\partial L}{\partial {v^a}_\mu}\xi_Y^\nu \mathbf{d}x^a \wedge \mathbf{d}^{n-1}X_{\mu\nu}, \tag{5.21}$$

where $\mathbf{d}^{n-1}X_{\mu\nu} = \mathbf{i}_{\partial/\partial\nu}\mathbf{d}^nX_\mu$. While it is of interest to consider general group actions, we are particularly interested here in those which are symmetries of the Lagrangian system. To make this precise, we say that the Lagrangian is ***equivariant*** with respect to the prolongation of the group action Φ if

$$L(g\cdot(X,x,v))\mathbf{d}^{n+1}X = L(X,x,v)(g_{\mathcal{X}}^{-1})^*(\mathbf{d}^{n+1}X).$$

In such cases we say that G is a ***symmetry*** of the Lagrangian.

Observe that equivariance of the Lagrangian is not the same as the Lagrangian being invariant under the prolonged group action. Invariance would simply mean that $L(g\cdot(X,x,v)) = L(X,x,v)$, and it turns out that this is not sufficient for $g\cdot\phi$ to be a solution whenever ϕ is. The reason that it is necessary to include the transformation of the volume form $\mathbf{d}^{n+1}X$ is that invariance of solutions (that is, solutions map to solutions) relies upon invariance of the action, and invariance of the action requires equivariance of the Lagrangian, as we will see explicitly below. This distinction is only

important if the symmetry action has non-zero base space components, such as a time scaling or reparameterization.

A necessary condition for the Lagrangian to be equivariant is infinitesimal equivariance, which is simply the derivative with respect to g of the definition of equivariance. That is, the Lagrangian is ***infinitesimally equivariant*** with respect to the prolonged group action if

$$\mathbf{d}L \cdot \xi_{J^1Y} = -L \operatorname{div}(\xi_{\mathcal{X}}).$$

This is simply the derivative of the above definition of a symmetry with respect to g in the direction ξ at the identity, and it has coordinate expression

$$\frac{\partial L}{\partial X^\mu}\xi_Y^\mu + \frac{\partial L}{\partial x^a}\xi_Y^a + \frac{\partial L}{\partial v^a{}_\mu}\left[\frac{\partial \xi_Y^a}{\partial x^b}v^b{}_\mu + \frac{\partial \xi_Y^a}{\partial X^\mu} - \frac{\partial \xi_Y^\nu}{\partial X^\mu}v^a{}_\nu\right] + L\frac{\partial \xi_Y^\mu}{\partial X^\mu} = 0.$$

We will now show that whenever the Lagrangian is equivariant under the prolonged group action, the corresponding momentum map is a conserved quantity.

Theorem 5.1 (Noether's theorem). *Consider a Lagrangian system $L : J^1Y \to \mathbb{R}$ which is equivariant under the prolongation of a left action $\Phi : G \times Y \to Y$ as described above. Then the corresponding Lagrangian momentum map $J_L : J^1Y \to \mathfrak{g}^* \otimes \Lambda^n(Y)$ given by (5.20) or (5.21) satisfies the global conservation law*

$$\int_{\partial\phi_{\mathcal{X}}(\mathcal{U}')} (j^1(\phi \circ \phi_{\mathcal{X}}^{-1}))^* J_L(\xi) = 0 \tag{5.22}$$

and the equivalent local conservation law

$$\mathbf{d}\Big[(j^1(\phi \circ \phi_{\mathcal{X}}^{-1}))^* J_L(\xi)\Big] = 0 \tag{5.23}$$

for all $\xi \in \mathfrak{g}$ and all subsets $\mathcal{U}'$ of $\mathcal{U}$.

Proof. The action of G on Y induces an action of G on the space of configurations $\mathcal{C}(Y)$ by pointwise action, so that $\Phi_g^{\mathcal{C}(Y)} : \mathcal{C}(Y) \to \mathcal{C}(Y)$ is given by $\Phi_g^{\mathcal{C}(Y)}(\phi)(U) = g(\phi(U))$. We now see that

equivariance of L implies

$$\begin{aligned} S'(g\cdot\phi) &= \int_{g_{\mathcal{X}}(\phi_{\mathcal{X}}(\mathcal{U}'))} L\Big(j^1(\phi'\circ(\phi'_{\mathcal{X}})^{-1})\Big)\,\mathbf{d}^{n+1}X \\ &= \int_{g_{\mathcal{X}}(\phi_{\mathcal{X}}(\mathcal{U}'))} L\Big(g\circ j^1(\phi\circ\phi_{\mathcal{X}}^{-1})\circ g_{\mathcal{X}}^{-1}\Big)\,\mathbf{d}^{n+1}X \\ &= \int_{\phi_{\mathcal{X}}(\mathcal{U}')} L(g\cdot j^1(\phi\circ\phi_{\mathcal{X}}^{-1}))(g_{\mathcal{X}})^*(\mathbf{d}^{n+1}X) \\ &= \int_{\phi_{\mathcal{X}}(\mathcal{U}')} L(j^1(\phi\circ\phi_{\mathcal{X}}^{-1}))\mathbf{d}^{n+1}X \\ &= S'(\phi) \end{aligned}$$

and so the action is invariant under the action of $\Phi^{\mathcal{C}(Y)}$. If ϕ is an extrema of the action, then invariance implies that $g\cdot\phi$ is also an extrema, and so the space of solutions is invariant under the group action. That is, $g\cdot\mathcal{C}_L(Y)=\mathcal{C}_L(Y)$.

If we now denote the infinitesimal generator of the group action on the space of configurations by $\xi_{\mathcal{C}(Y)}:\mathcal{C}(Y)\to T(\mathcal{C}(Y))$, then invariance of the action can be written $\mathbf{d}S'(\phi)\cdot\xi_{\mathcal{C}(Y)}=0$ for all $\xi\in\mathfrak{g}$, which still holds if we restrict to the space of solutions $\mathcal{C}_L(Y)$.

Using (5.16), however, we can also write the derivative of the action in the group direction as

$$\mathbf{d}S'(\phi)\cdot\xi_{\mathcal{C}(Y)} = \int_{\partial\phi_{\mathcal{X}}(\mathcal{U}')} (j^1(\phi\circ\phi_{\mathcal{X}}^{-1}))^*\Big(\mathbf{i}_{\xi_{J^1Y}}\Theta_L\Big),$$

where we have used the fact that $\xi_{J^1Y}=j^1\xi_Y$. Using the definition of the Lagrangian momentum map and the above statement of invariance of the action, we now have

$$\int_{\partial\phi_{\mathcal{X}}(\mathcal{U}')} (j^1(\phi\circ\phi_{\mathcal{X}}^{-1}))^* J_L(\xi) = \mathbf{d}S'(\phi)\cdot\xi_{\mathcal{C}(Y)} = 0,$$

which is the global statement of Noether's theorem.

Applying Stokes' theorem shows that this is equivalent to

$$\int_{\phi_{\mathcal{X}}(\mathcal{U}')} \mathbf{d}\Big[(j^1(\phi\circ\phi_{\mathcal{X}}^{-1}))^* J_L(\xi)\Big] = 0$$

for any $\mathcal{U}'\subset\mathcal{U}$, and thus we can conclude that the integrand itself is zero, giving the local (or divergence) statement of Noether's theorem. □

The above proof shows that in fact only infinitesimal equivariance is required for Noether's theorem, rather that the stronger statement of equivariance itself. This is often useful in examples.

In the above theorem we have not explicitly accounted for boundary conditions, and the assumption of equivariance requires that body forces arising from external potentials in the Lagrangian

do not act in the symmetry direction. If we now consider a more general situation, in which the solution satisfies traction boundary conditions in the sense of equation (5.11) and we do not require equivariance of the Lagrangian, then for an arbitrary variation V we have

$$\mathbf{d}S(\phi)\cdot V = \int_{\phi_{\mathcal{X}}(\partial_1\mathcal{U})} \tau\cdot V\ dA + \int_{\phi_{\mathcal{X}}(\partial\mathcal{U}\setminus\partial_1\mathcal{U})} (j^1(\phi\circ\phi_{\mathcal{X}}^{-1}))^*\Big(\mathbf{i}_{j^1V}\Theta_L\Big)$$

and so taking the variation to be $V=\xi_{\mathcal{C}(Y)}$ gives us

$$\int_{\phi_{\mathcal{X}}(\partial\mathcal{U}\setminus\partial_1\mathcal{U})} (j^1(\phi\circ\phi_{\mathcal{X}}^{-1}))^* J_L(\xi) = -\int_{\phi_{\mathcal{X}}(\partial_1\mathcal{U})} \tau\cdot\xi_Y\ dA + \mathbf{d}S(\phi)\cdot\xi_{\mathcal{C}(Y)}. \tag{5.24}$$

5.2.6 Symmetries and momentum maps

We now turn to considering the three main symmetries which arise in continuum solid mechanics problems. These are translation, rotation and time translation invariance, and they give rise to conservation of linear momentum, angular momentum and energy, respectively.

Translation invariance. The group of translations is $\mathcal{G}^{\text{tr}}\cong Y_X=\mathbb{R}^3$, and it acts by $\eta^r(X,x)=(X,x+r)$. The infinitesimal generator corresponding to $\xi^r\in\mathfrak{g}^{\text{tr}}$ is thus given by $\xi^r(X,x)=(X,x,0,r)$ for each $r\in\mathbb{R}^3$.

The Lagrangian (5.1) is clearly equivariant because it has no explicit dependence on the fiber spatial coordinate x. Computing the Lagrangian momentum map gives

$$J_L(\xi^r)=\frac{\partial L}{\partial {v^a}_\mu} r^a\ \mathbf{d}^nX_\mu$$

and it can be easily seen that the local Noether's theorem (5.23) recovers the Euler-Lagrange equations[1].

Using the global form of Noether's theorem with boundary conditions (5.24) and assuming that $\phi_{\mathcal{X}}=\text{id}$, we compute the various terms to be

$$\begin{aligned}
\int_{\partial\mathcal{X}\setminus\partial_1\mathcal{X}} (j^1(\phi\circ\phi_{\mathcal{X}}^{-1}))^* J_L(\xi) &= \int_{\mathcal{B}} \frac{\partial L}{\partial {v^a}_t}\bigg|_{t=T}\ dV + \int_0^T\int_{\partial_d\mathcal{B}} \frac{\partial L}{\partial {v^a}_i} N_i\ dAdt\\
\int_{\partial_1\mathcal{X}} \tau\cdot\xi_Y\ dA &= \int_0^T\int_{\partial_\tau\mathcal{B}} \tau_a\ dAdt + \int_{\mathcal{B}} (p_0)_a(-dV)\\
\mathbf{d}S(\phi)\cdot\xi_{\mathcal{C}(Y)} &= \int_0^T\int_{\mathcal{B}} \frac{\partial L}{\partial x^a}\ dVdt.
\end{aligned}$$

If we now substitute in the Lagrangian (5.1) and use the expressions (5.3), then we see that (5.24)

[1] This can also be predicted from general theory, because the action of $\mathcal{G}^{\text{tr}}$ is vertically transitive (see Gotay et al. [1997])

becomes

$$\int_{\mathcal{B}} p_a(T,X)\ dV - \int_{\mathcal{B}} p_a(0,X)\ dV = \int_0^T \int_{\partial_d \mathcal{B}} P_a{}^i(t,X) N_i(X)\ dA dt + \int_0^T \int_{\partial_\tau \mathcal{B}} T_a(X)\ dA dt + \int_0^T \int_{\mathcal{B}} B_a(t,x)\ dV dt.$$

This shows how the whole-body linear momentum changes from time 0 to time T under the influence of traction boundary forces $T_a = -\tau_a$, displacement boundary conditions, and body forces $B_a = -\nabla_a V$. In the case of free boundary conditions, when $\partial_\tau \mathcal{B} = \mathcal{B}$ and $\tau = 0$, and zero body forces, we recover the conservation of whole-body linear momentum.

Rotation invariance. The group of rotations is $\mathcal{G}^{\text{rot}} \cong \text{SO}(3)$, with action given by $\eta^R(X,x) = (X, \exp(R)x)$ for each skew matrix $R \in \mathfrak{so}(3)$. The infinitesimal generator for an element $\xi^R \in \mathfrak{g}^{\text{rot}}$ is given by $\xi^R(X,x) = (X,x,0,Rx)$.

The assumption of material frame indifference, namely that the stored energy function W in (5.1) depends only on $F^T F$, means that the Lagrangian itself is invariant under the action of $\mathcal{G}^{\text{rot}}$. The Lagrangian momentum map is

$$J_L(\xi^R) = \frac{\partial L}{\partial v^a{}_\mu} R^a{}_b \varphi^b\ \mathbf{d}^n X_\mu$$

and so the local Noether's theorem is the statement that $R_{ab}\sigma^{ba} = 0$ for all R, so skew-symmetry of R implies that the Cauchy stress tensor σ is symmetric. This recovers the standard balance of moment of momentum. The global Noether's theorem is simply the statement of global angular momentum conservation, assuming compatible boundary conditions.

Time translation invariance. The group of time translations is $\mathcal{G}^{\text{time}} \cong \mathbb{R}$, with action $\eta^\alpha(t,X,x) = (t+\alpha, X, x)$ and $\xi^\alpha \in \mathfrak{g}^{\text{time}}$ for each $\alpha \in \mathbb{R}$. The infinitesimal generator for $\xi^\alpha \in \mathfrak{g}^{\text{time}}$ is $\xi^\alpha(t,X,x) = (t,X,x,\alpha,0,0)$. The Lagrangian (5.1) is equivariant with respect to the action of $\mathcal{G}^{\text{time}}$ as it is independent of time and the Lagrangian momentum map gives

$$(j^1(\phi \circ \phi_{\mathcal{X}}^{-1}))^* J_L(\xi^\alpha) = \left[-\frac{\partial L}{\partial v^a{}_\mu} \varphi^a{}_{,t}\ \mathbf{d}^n X_\mu - E_L \mathbf{d}^n X_t \right] \alpha.$$

The local Noether's theorem then gives the local energy continuity equation, while the global Noether's theorem gives the statement of whole-body energy conservation. In fact, arbitrary time reparameterizations are a symmetry of the system, and also lead to energy conservation. In considering such actions it is crucial to distinguish between equivariance and invariance of the Lagrangian.

Chapter 6

Multisymplectic asynchronous variational integrators

In this chapter we wish to consider both the abstract theory of multisymplectic discretizations, as well as the particular example of Asynchronous Variational Integators (AVI). We begin with the concrete example of AVI in §6.1, and then in §6.2 and §6.3 we consider the abstract theory.

6.1 Asynchronous variational integrators (AVIs)

In this section we describe a class of asynchronous time integrators. This work has appeared in Lew et al. [2003a] and Lew et al. [2003b].

6.1.1 Systems of particles

Consider a system of N_a particles with dynamics described by a Lagrangian of the form

$$L(x, \dot{x}) = \sum_a \frac{1}{2} m_a \|\dot{x}_a\|^2 - \sum_K V_K(x), \tag{6.1}$$

where m_a is the mass of particle a at position $x_a \in \mathbb{R}^3$ and V_K is the potential energy of subsystem K. It is this decomposition of the total potential energy which permits the asynchronicity of the time discretization.

Example: Molecular dynamics. One common example of systems of the form (6.1) arises in molecular problems. Here the potentials V_K represent the interaction between pairs (or triples) of particles, so that we have $V_{ab} = V(\|x_b - x_a\|)$ for all $a < b$ and some given $V(r)$.

Example: Finite elements. Continuum mechanical systems can be specified by considering the deformation mapping $\varphi : \mathcal{B} \to \mathbb{R}^3$ which acts on points X in the reference configuration $\mathcal{B} \subset \mathbb{R}^3$. The deformation of local or infinitesimal neighborhoods is described by the deformation gradient

$F = \nabla_X \varphi$, where $\nabla_X \varphi$ denotes only the spatial gradient of φ. The time derivative of φ is denoted with $\dot{\varphi}$, while accelerations are indicated with $\ddot{\varphi}$. The Lagrangian is of the form[1]

$$L(\varphi, \dot{\varphi}) = \int_{\mathcal{B}} \left(\frac{\rho_0}{2} \|\dot{\varphi}\|^2 - W(\nabla_X \varphi, X) \right) dX, \tag{6.2}$$

which describes the dynamics of hyperelastic materials. In equation (6.2) ρ_0 is the density of the continuum in the reference configuration and $W(F, X)$ is the free-energy density of the material. We assume that $W(F, X)$ satisfies the requirement of material frame indifference (see, e.g., Marsden and Hughes [1994]). Recall that for hyperelastic materials the first Piola-Kirchhoff stress tensor is given by

$$P_I^i = \frac{\partial}{\partial F_I^i} W(F, X). \tag{6.3}$$

The Euler-Lagrange equations for (6.2) translate into the infinitesimal balance of linear momentum

$$\nabla_X \cdot P - \rho_0 \ddot{\varphi} = 0. \tag{6.4}$$

Consider now the spatial semidiscretization of (6.2) by finite elements. We let $\mathcal{T}$ be a triangulation of $\mathcal{B}$. The corresponding finite-dimensional space of finite-element solutions consists of deformation mappings of the form

$$\varphi_h(X) = \sum_{a \in \mathcal{T}} x_a \, N_a(X), \tag{6.5}$$

where N_a is the shape function corresponding to node a, x_a represents the position of the node in the deformed configuration.

We let m_a be the lumped mass associated to node a arising from the discretization (6.5) and some lumping scheme. The elemental potential energies $V_K(x_K)$ are given by

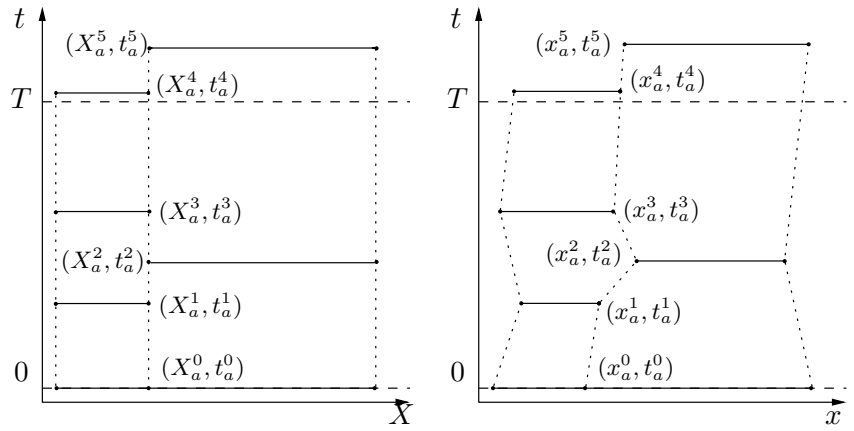
$$V_K(x_K) = \int_K W(\nabla \varphi_h, X) dX, \tag{6.6}$$

where x_K is the vector of positions of all the nodes in element K. Of course, terms in the potential energy resulting from boundary conditions or body forces should be added to the expression in (6.6). We thus have a finite dimensional system of the form (6.1).

6.1.2 Asynchronous time discretizations

While asynchronous methods can be applied to any Lagrangian system with a decomposed potential energy, here we use the terminology associated with finite element discretizations. To apply the

[1] Of course terms representing body forces arising from a potential energy could be added to the Lagrangian density (6.2). Similarly, the Lagrangian density could depend explicity on time. This cases do not present any essential difficulties and can be found in Lew et al. [2003a]. For simplicity, we will not consider them here.



(a) Reference configuration (b) Deformed configuration

Figure 6.1: Spacetime diagram of the motion of a two-element, one-dimensional mesh. The set of coordinates and times for a single node is shown in the reference and deformed configuration. Note that the nodal coordinates and times are labeled according to the interaction of the node with all elements to which it belongs.

methods to a problem such as molecular dynamics, simply think of "elements" as being "pairwise potentials", and "nodes" as being "particles".

Time discretization. For each element K we choose a time step Δt_K. Different elements could have different time steps[2]. We denote with $t_K^j = j\Delta t_K$, $j = 0, \ldots, N_K$, the time of the j-th time step for element K, with N_K being the smallest integer such that $t_K^j \geq T$. From this, we construct the space-time discretization as shown in Figure 6.1. As shown in Figure 6.1, we also denote with t_a^i, $i = 0, \ldots, N_a$, the time instant for node a. Nonconstant time step per element are considered in Lew et al. [2003a].

Discrete action sum. Define the discrete action sum to be

$$S_d = \sum_a \sum_{i=0}^{N_a - 1} \frac{1}{2} m_a (t_a^{i+1} - t_a^i) \left\| \frac{x_a^{i+1} - x_a^i}{t_a^{i+1} - t_a^i} \right\|^2 - \sum_K \sum_{j=0}^{N_K - 1} (t_K^{j+1} - t_K^j)\, V_K(x_K^{j+1}), \tag{6.7}$$

which approximates the continuous action over the time interval $[0, T]$.

Discrete Lagrangians. In the particular case of finite element discretizations, the above action sum can be realized as the sum of space-time discrete Lagrangians. Let $m_{K,a}$ be the mass of node

[2]It is straightforward to extend the definition of AVI methods given here to allow each elemental time step to vary as the integration advances in time, either in response to error driven adaptivity criteria, or to some other information such as local energy balance laws (see Lew et al. [2003a]).

a which is due to element K, so that $m_a = \sum_K m_{K,a}$. Then the discrete action is the sum over all elements K and elemental times j of the discrete Lagrangian

$$L_K^j = \sum_{a\in K} \sum_{\{i:\, t_K^j \le t_a^i < t_K^{j+1}\}} \frac{1}{2} m_{K,a}(t_a^{i+1} - t_a^i) \left\| \frac{x_a^{i+1} - x_a^i}{t_a^{i+1} - t_a^i} \right\|^2 - (t_K^{j+1} - t_K^j)\, V_K(x_K^{j+1}), \tag{6.8}$$

where x_K is the vector of positions of all the nodes in element K. The discrete Lagrangian L_K^j approximates the incremental action of element K over the interval $[t_K^j, t_K^{j+1}]$. In general, L_K^j depends on the nodal positions x_a^i, $a \in K$, and such that $t_a^i \in [t_K^j, t_K^{j+1}]$[3]. The discrete action sum S_d thus depends on all the nodal positions x_a^i for all nodes $a \in \mathcal{T}$ and for all times t_a^i, $0 \le i \le N_a$.

Discrete Euler-Lagrange equations. The discrete variational principle states that the discrete trajectory of the system should be a critical point of the action sum for all admissible variations of the nodal coordinates x_a^i. The discrete Euler-Lagrange equations are

$$D_a^i S_d = 0 \tag{6.9}$$

for all $a \in \mathcal{T} \setminus \partial_d \mathcal{B}$ such that $t_a^i \in (0, T)$. Here and subsequently, D_a^i denotes differentiation with respect to x_a^i. The discrete Euler-Lagrange equations (6.9) define the equations of motion of the discrete problem.

After introducing the definition of the discrete Lagrangian (6.8) under consideration into (6.9), a straightforward calculation gives the discrete Euler-Lagrange equations explicitly in the form

$$p_a^{i+1/2} - p_a^{i-1/2} = I_a^i, \tag{6.10}$$

where

$$p_a^{i+1/2} \equiv m_a \frac{x_a^{i+1} - x_a^i}{t_a^{i+1} - t_a^i} \equiv m_a\, v_a^{i+1/2} \tag{6.11}$$

are discrete linear momenta and m_a are the nodal masses, i.e., $m_a = \sum_K m_{K,a}$. In addition, we define

$$I_K^j \equiv -(t_K^j - t_K^{j-1}) \frac{\partial}{\partial x_K^j} V_K(x_K^j), \tag{6.12}$$

which may be regarded as the impulses exerted by element K on its nodes at time t_K^j. In equation (6.10) I_a^i represents the component of I_K^j corresponding to node a, with $t_a^i = t_K^j$. Eq. (6.10) may be interpreted as describing a sequence of percussions imparted by the elements on their nodes at discrete instants of time. Thus, the element K accumulates and stores impulses I_K^j over the time interval (t_K^{j-1}, t_K^j). At the end of the interval, the element releases its stored impulses by imparting

[3] If adjacent elements possess coincident elemental times one must derive the correct interpretation by taking the appropriate limits. See [Lew et al., 2003a, §3.2]

percussions on its nodes, causing the linear momentum of the nodes to be altered. The resulting nodal trajectories can be regarded as piecewise linear in time. We note that adjacent elements interact by transferring linear momentum through their common nodes. Note that the resulting algorithm is explicit of the central-difference type.

6.1.3 Implementation of AVIs

In this section we turn our attention to discussing the implementation of the AVI corresponding to the discrete Lagrangian (6.8).

Because of the algorithm's asynchronous nature, a suitable scheduling procedure which determines the order of operations while ensuring causality must be carefully designed. One particularly efficient implementation consists of maintaining a priority queue (see, e. g., Knuth [1998]) containing the elements of the triangulation.[4] The elements in the priority queue are ordered according to the next time at which they are to become active. Thus, the top element in the queue, and consequently the next element to be processed, is the element whose next activation time is closest to the present time.

The general flow of the calculations is as follows. The priority queue is popped in order to determine the next element to be processed. The new configuration of this active element is computed from the current velocities of the nodes. Subsequently, these velocities are modified by impulses computed based on the new element configuration. Finally, the next activation time for the element is computed and the element is pushed into the queue. A flow chart of the numerical procedure is given in fig. 6.2. Note that the algorithm allows the time step of each element to change in time.

The use of the priority queues is particularly simple in C++, since these are provided by several freely available libraries. Also, routines implementing priority queues in C are freely available to download. Priority queues are frequently implemented through balanced binary trees [Knuth, 1998, pp.458].

The adaption of existing FE codes to implement the explicit AVI integrator is fairly simple. The computations at the element level remain untouched, while the driver that assembles the global internal force vector should be removed, since there is no assembly required. Instead, a driver that implements the operations in fig. 6.2 should be coded. Notice that apart from the priority queue and two arrays to store elemental and nodal times respectively, no extra storage nor new arrays are required over traditional explicit Newmark schemes. To plot the configuration of the continuum, a short routine computing the positions of the nodes at the time of the most advanced element is needed. In this case, each node is advanced by following a linear trajectory parallel to its velocity. It is noteworthy that explicit AVIs allows for the reusage of most of the existing FE structural

[4] We stress that this is only one possibility for ordering the computation, and is not imposed by the discrete Euler-Lagrange equations.

Explicit AVI Algorithm

Input data: $T, \mathcal{T}, \left\{ x_a^0, \dot{x}_a^{1/2} \mid a \in \mathcal{T} \right\}$

- ▷ Initialization
  - ▷ $q_a \leftarrow x_a^0$, $v_a \leftarrow \dot{x}_a^{1/2}$, $\tau_a \leftarrow 0$ for all $a \in \mathcal{T}$
  - ▷ Do for all $K \in \mathcal{T}$
    - ▷ $\tau_K \leftarrow 0$
    - ▷ Compute t_K^1
    - ▷ Push (t_K^1, K) into priority queue
  - ▷ End Do
- ▷ Iterate over the elements in time
  - ▷ Do until priority queue is empty
    - ▷ Extract next element: Pop (t, K) from priority queue
    - ▷ Update positions: $q_a \leftarrow q_a + v_a\,(t - \tau_a)$, for all $a \in K$
    - ▷ Update node's time: $\tau_a \leftarrow t$, for all $a \in K$
    - ▷ If $t < T$
      - ▷ Update velocities:
        $v_a \leftarrow v_a - (t - \tau_K)\,\frac{\partial V_K}{\partial x_a}\,(q_K)\,/m_a$, for all $a \in K$
      - ▷ Update element's time: $\tau_K \leftarrow t$
      - ▷ Compute t_K^{next}
      - ▷ Schedule K for next update:
        Push (t_K^{next}, K) into priority queue
  - ▷ End Do
- ▷ End

Figure 6.2: Algorithm implementing the discrete Euler-Lagrange equations of the action sum given by equation (6.8).

dynamics codes.

6.1.4 Momentum conservation properties

In order to derive the discrete momentum balance equations, we consider a one-parameter group of trajectories with nodal positions given by $(x_a^i)^\varepsilon = \exp(\varepsilon\Omega)x_a^i + \varepsilon v$, for any vector $v \in \mathbb{R}^3$ and skew-symmetric matrix Ω. We assume that the discrete action sum is translation and rotation invariant, so that the value of S_d for $(x_a^i)^\varepsilon$ is equal to its value for x_a^i.

Differentiating the discrete action sum of the perturbed trajectory with respect to ε gives

$$0 = \frac{\partial}{\partial\varepsilon} S_d\Big((x_a^i)^\varepsilon\Big)\Big|_{\varepsilon=0} = \sum_a \sum_{i=0}^{N_a} D_a^i S_d \cdot v + \sum_a \sum_{i=0}^{N_a} (x_a^i \times D_a^i S_d) \cdot \omega \tag{6.13}$$

where $\omega \in \mathbb{R}^3$ is the axial vector of Ω, which in terms of the Hodge star operator is $\omega = *\Omega$.

We now use the fact that x_a^i is a trajectory of the system and so it satisfies the discrete Euler-Lagrange equations (6.10). Then $D_a^i S_d = 0$ for all a and $i = 1, \ldots, (N_a - 1)$. Under these conditions, equation (6.13) reduces to

$$0 = \left[\sum_a D_a^0 S_d + \sum_a D_a^{N_a} S_d\right] \cdot v + \left[\sum_a x_a^0 \times D_a^0 S_d + \sum_a x_a^{N_a} \times D_a^{N_a} S_d\right] \cdot \omega. \tag{6.14}$$

As this must hold for all v and all ω we obtain

$$\sum_a D_a^{N_a} S_d = -\sum_a D_a^0 S_d \tag{6.15}$$

$$\sum_a x_a^{N_a} D_a^{N_a} S_d = -\sum_a x_a^0 D_a^0 S_d, \tag{6.16}$$

which furnishes a precise statement of discrete linear and angular momentum conservation, respectively. Using the particular expression (6.7) for the discrete action sum we can evaluate (6.15) and (6.16) to give

$$\sum_a p_a^{1/2} = \sum_a p_a^{N_a - 1/2} \tag{6.17}$$

$$\sum_a x_a^0 \times p_a^{1/2} = \sum_a x_a^{N_a} \times p_a^{N_a - 1/2}, \tag{6.18}$$

which clearly shows that both momenta are conserved.

Momentum evolution without symmetry. We assumed above that the discrete action sum was translation and rotation invariant, which will only be the case if there are no displacement boundary conditions and no forces or torques are applied to the body. If there are such effects, then

we can carry out the same calculation but with the left-hand side of (6.13) being nonzero. In fact, this nonzero term will exactly give the net resultant force and moment and so (6.17) and (6.18) will give discrete versions of the familiar statements that change in linear and angular momentum should equal the total impulse and moment of impulse imparted to the system. These expressions are worked out in general in Lew et al. [2003a].

6.1.5 Numerical examples

In this section we illustrate the performance of an explicit AVI for structural dynamics by simulating the motion in vacuum of a simplified model of the blades of an Apache AH-64 helicopter. These and other simulations in this section were performed using code written by Adrian Lew.

The simulation of the dynamics of rotor blades and similar systems has long been a challenge and a test-bed for time-integration algorithms for structural dynamics (Borri [1986]; Friedmann [1990]; Sheng, Fung, and Fan [1998]; Armero and Romero [2001a,b]; Bottasso and Bauchau [2001]). The slender geometry of the blades renders the system prone to dynamic instability, and the need to compute accurate solutions for long periods of time, to conserve linear and angular momentum, and to accurately distribute the energy across the frequency spectrum place stringent tests on time-integration methods.

The blade has a span of 7.2 m, a chord of 533 mm and a maximun thickness of 40 mm. The three-dimensional mesh of the blade is shown in Figure 6.3. It consists of 2089 ten-noded tetrahedral elements and 4420 nodes. The original geometry of the hub of the blade rotor was replaced by a simple straight joint. The blades comprise 5 steel spars reinforced with glass fiber. However, for this example we simply replace the hollow structure by an effective homogeneous material. The strain-energy density is given by

$$W(F,X) = \frac{\lambda_0(X)}{2}\left(\log J\right)^2 - \mu_0(X)\log J + \frac{\mu_0(X)}{2}\mathrm{tr}\left(F^{\mathsf{T}}F\right), \tag{6.19}$$

which describes a Neohookean solid extended to the compressible range. In this expression, $\lambda_0(X)$ and $\mu_0(X)$ are—possibly inhomogeneous—Lamé constants. The corresponding Piola stress-deformation relation follows from (6.3) in the form:

$$P = \lambda_0 \log J\, F^{-\mathsf{T}} + \mu_0\left(F - F^{-\mathsf{T}}\right). \tag{6.20}$$

As the initial condition at time $t = 0$, we consider the blade in its undeformed configuration with a velocity field corresponding to an angular velocity $\omega = 40$ rad/s. Due to the absence of external forces or dissipation the energy is conserved and the center of mass does not move from its original

position for all times. We classify cases according to the dimensionless angular velocity

$$\hat{\omega} = \frac{\omega L^2}{cw},$$

where ω is the nominal angular velocity of the blade, L its span, w is the chord of the blade, and

$$c = \sqrt{\frac{\lambda_0 + 2\mu_0}{\rho_0}},$$

is a nominal dilatational wave speed of the material. In all examples, the elemental time steps are determined from the Courant condition, which provides an estimate of the stability limit for explicit integration (see, e.g., Hughes [1987]). Specifically, the value of the time step for each element is set to a fraction of the Courant limit and is computed as

$$\Delta t = f \, \frac{h}{c}, \tag{6.21}$$

where $f = 1/10$ and h is the radius of the largest ball contained in the element. The time step is kept constant in each element throughout the computation. We note that the elemental time steps are not required to be integer-related, and the element trajectories are not synchronized in general.

We proceed to report three test cases which illustrate the performance and properties of the AVIs. The first case consists of a homogeneous blade with material constants $\rho_0 = 250$ kg/m^3, $\lambda_0 = 0.98$ GPa and $\mu_0 = 0.98$ GPa, and, correspondingly, $\hat{\omega} = 1.13$. For this choice of parameters the computed trajectory is close to a rigid rotation, Fig. 6.4, and remains stable both for short (Fig. 6.4(a)) and long (Fig. 6.4(b)) times.

The second case consists of a homogeneous blade with material constants $\rho_0 = 2500$ kg/m^3, $\lambda_0 = 4$ GPa and $\mu_0 = 4$ GPa, and, correspondingly, $\hat{\omega} = 1.78$. Moderately large amplitude oscillations, including spanwise torsional modes, develop during the initial transient, Fig. 6.5(a). Remarkably, after this transient the blade settles down to a fairly stable shape and rotates almost rigidly.

Finally we consider a two-material blade. The central joint consists of a stiff material with constants $\rho_0 = 4500$ kg/m^3, $\lambda_0 = 3$ GPa and $\mu_0 = 0.75$ GPa, while the remainder of the blade is composed of a softer material with constants $\rho_0 = 250$ kg/m^3, $\lambda_0 = 0.1$ GPa and $\mu_0 = 0.025$ GPa. The contrast between the dilatational wave speeds of the two materials, and therefore between the stable time steps for elements of the same size in the two different regions, is 1.3. The value of the dimensionless angular velocity velocity relative to the softer material is $\hat{\omega} = 5.02$. The initial transient is characterized by very large-amplitude oscillations in the blade, Fig. 6.6(a), while the central joint remains essentially rigid. As in the previous case, after this transient the blade settles down to a stable nearly-rigid orbit.

A first noteworthy feature of the solutions is that, despite their asynchronous character, they

advance smoothly in time without ostensible jerkiness or vacillation. Note also that the center of mass of the blade does not move, which is a consequence of the conservation of linear momentum by the algorithm. The period of rotation for long times is of particular interest. A perfectly rigid blade would rotate once every 0.157 s. However, a flexible blade increases its span under the action of the inertia forces induced by the rotatory motion, and its angular velocity slows down in order to conserve the total angular momentum. This effect is indeed observed in the simulations, as shown in table 6.1, and is a consequence of angular-momentum conservation.

A measure of performance of the AVIs is shown in Fig. 6.7, which depicts the number of updates in each of the elements of the mesh for the third numerical example described above. As is evident from the figure, the larger elements in the mesh are updated much less frequently than the finer elements. In particular, a small number of slivers are updated very frequently, as required by their small Courant limit. Also, the elements in the central joint, made of a stiffer material, are updated more often than the more flexible elements in the blade.

Some relevant statistics are collected in Table 6.2. Overall, in the present example the number of AVI updates is roughly 15% of the number of updates required by explicit Newmark at constant time step. It should be carefully noted, however, that in the example under consideration the vast majority of the elements have similar sizes and aspect ratios, differing only by at most one order of magnitude. It is easy to set up examples in which the update count of the constant time step algorithm bears an arbitrarily large ratio to the update count of the AVI. A case which arises in practice with some frequency concerns a roughly uniform triangulation of the domain which contains a small number of high aspect-ratio elements. The presence of a single bad element suffices to drive down the critical time step for explicit integration to an arbitrarily small value. This problem often besets explicit dynamics, especially in three dimensions where bad elements, or *slivers*, are difficult to eliminate entirely. The AVI algorithm effectively sidesteps this difficulty, as bad elements drive down their own times steps only, and not the time steps of the remaining elements in the mesh. In this manner, the overall calculation is shielded from the tyranny of the errant few.

Case	$\hat{\omega}$	Period [s]
1	1.13	0.1575 ± 0.0005
2	1.78	0.1595 ± 0.0005
3	5.02	0.1845 ± 0.0005

Table 6.1: Period of rotation of the blade for long times. As the value of $\hat{\omega}$ grows, the blade deforms more increasing its span, and therefore its moment of inertia, to accommodate the centrifugal forces. Since the discrete angular momentum is conserved, the period rotation should grow accordingly.

In section §1.6.5 we alluded to the excellent energy-conservation properties of variational integrators. Our numerical tests (see also Lew et al. [2003a]) suggest that the AVIs possess excellent

Case	$\hat{\omega}$	Final time [s]	Maximum	Minimum	Total	Speed-Up
1	1.13	7.849	266×10^6	14×10^6	8.7×10^{10}	6.37
2	1.78	15.015	325×10^6	17×10^6	42.6×10^{10}	6.37
3	5.02	27.439	235×10^6	12×10^6	8.4×10^{10}	5.80

Table 6.2: Maximum and minimum number of elemental updates for a single element at the final time. The total column shows the sum of the number of elemental updates in the whole mesh at the final time. In contrast, traditional time stepping algorithms would have advanced with the same number of updates on each element, which is equal to the value in the Maximum column. The ratio between the total number of updates in the whole mesh in these two cases is shown in the Speed-up column, a direct measure of the cost saving features of AVI.

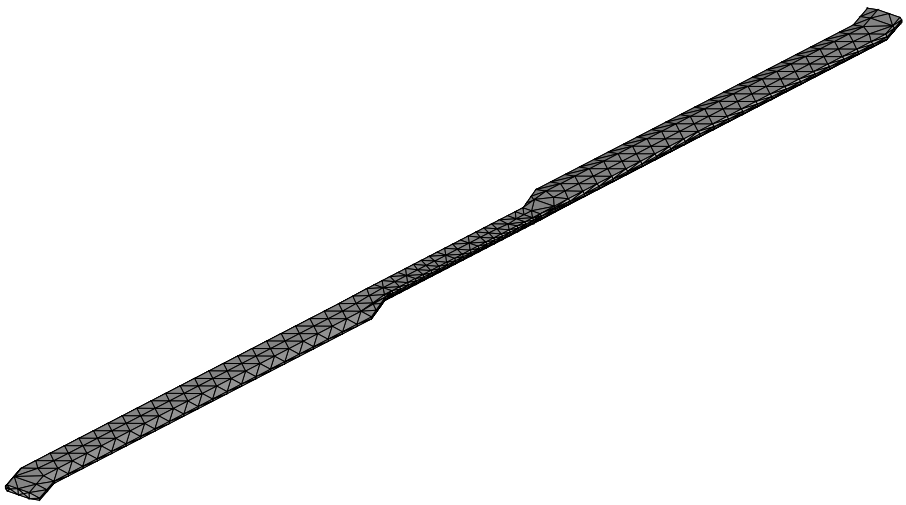
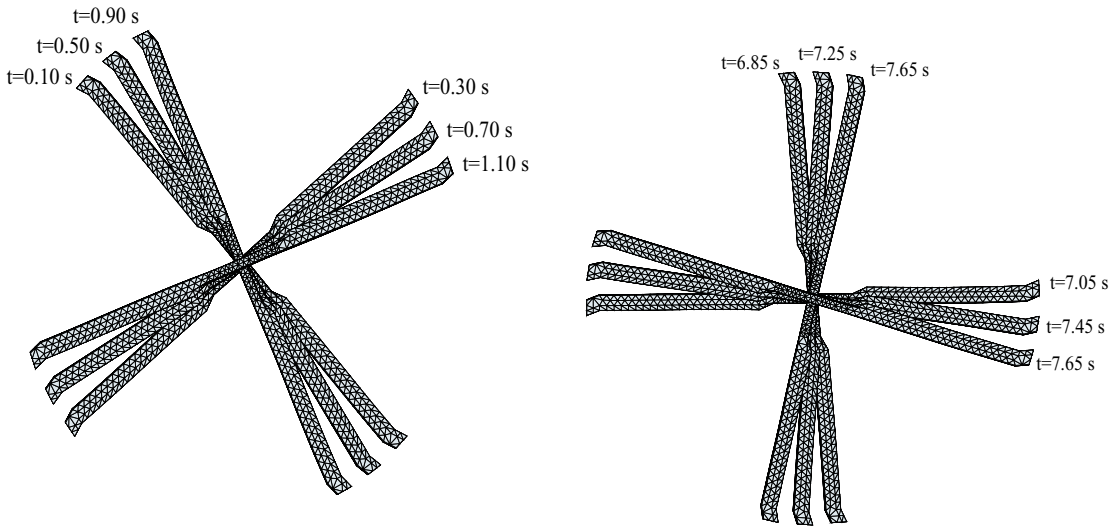


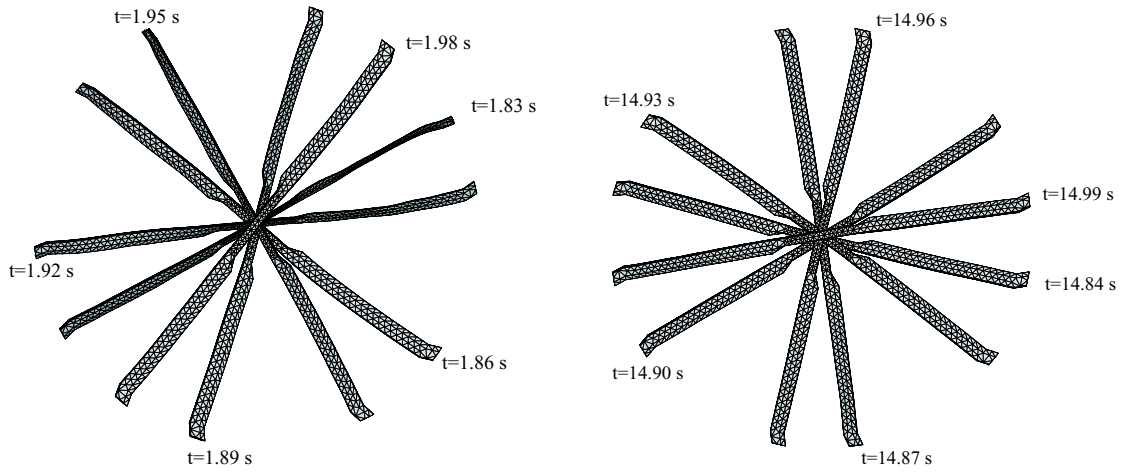
Figure 6.3: Mesh of the blade. It consists of 2089 ten-noded tetrahedral elements and 4420 nodes.



(a) Initial steps

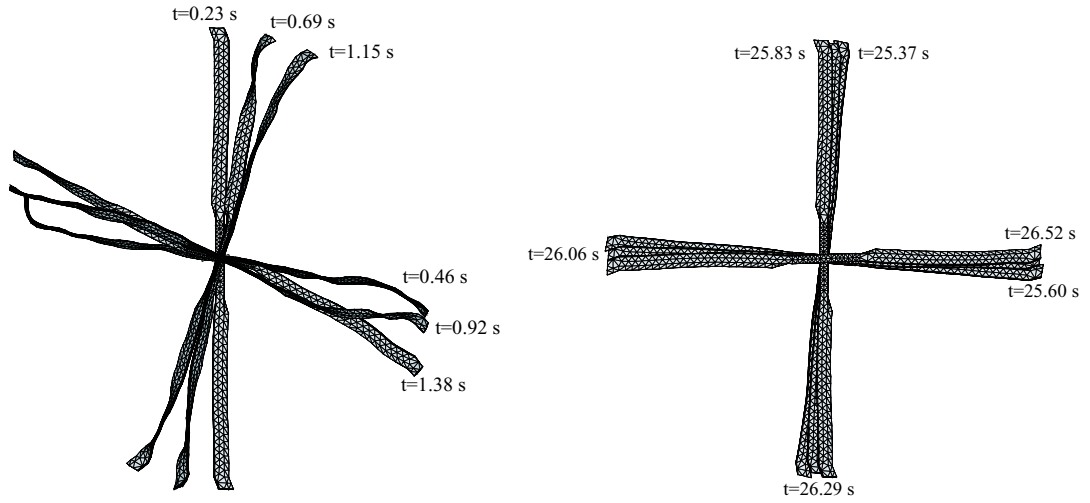
(b) Final steps

Figure 6.4: Evolution of the blade for the first and most rigid case. The motion of the blade is essentially that of a rigid body. The center of mass does not move, a consequence of the discrete linear momentum conservation, and the period of the blade is very close to the one of a completely rigid blade, since the spanwise elongation is negligible. The final snapshots correspond to approximately 266 million updates of the smallest element in the mesh.



(a) Initial steps (b) Final steps

Figure 6.5: Evolution of the blade for the second case. During the initial phases of the motion, some fairly large deflections, including torsion along the spanwise direction, occur. However, after a relatively long time the blade rotates with an almost fixed shape. The period of rotation has changed slightly with respect to a rigid blade, since there is a non-negligible spanwise elongation inducing a change in the corresponding moment of inertia. The final snapshots correspond to approximately 325 million updates of the smallest element in the mesh.



(a) Initial steps (b) Final steps

Figure 6.6: Evolution of the blade for the third and softest case. During the initial phases of the motion, the blade behaves as a very flexible strip. Surprisingly, after a relatively long time the blade settles to rotate with small amplitude oscillation close to an almost fixed shape. The period of rotation with respect to a rigid blade has changed considerably, since the spanwise elongation is large. The final snapshots correspond to approximately 234 million updates of the smallest element in the mesh.

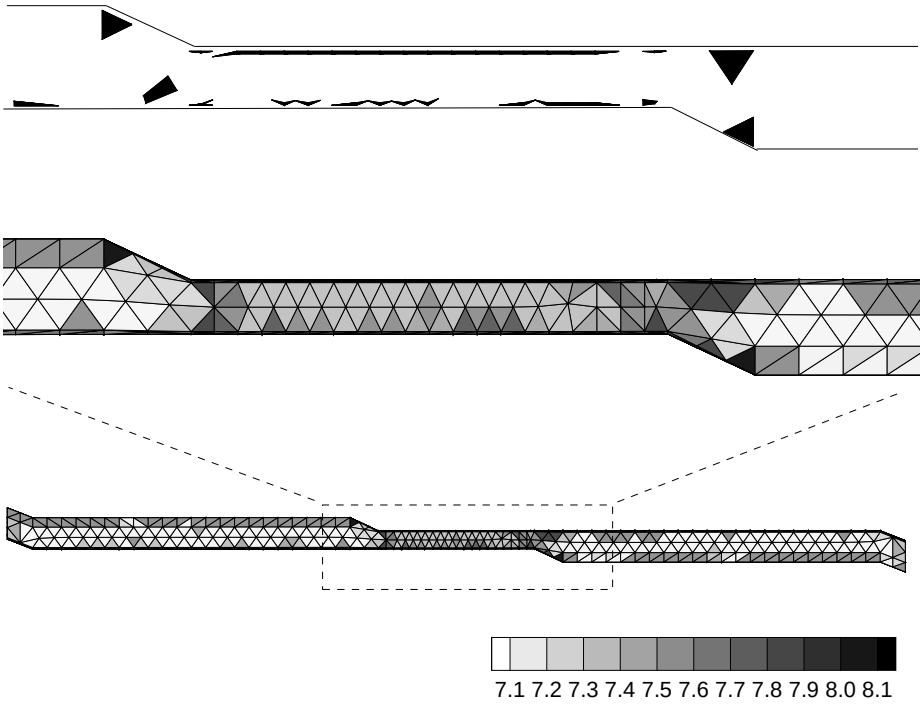


Figure 6.7: Contour plot of the $\log_{10}$ of the number of times each element was updated by the AVI after 27.439 s of simulation of case 3, in which inertial forces prevail. The picture on the middle shows an enlargement of the central part of the blade, which is made out of a stiffer material than the rest. The abrupt change in the number of elemental updates between the two region is noteworthy. Additionally, the picture on the top shows only those few elements updated more than 10^8 times. These elements or slivers would drive the whole computation down for a constant time step algorithm, while AVI circumvents this difficulty gracefully.

energy-conservation properties as well. Thus, for instance, Figure 6.8 shows the time evolution of the total energy of the blade in the third case. It is remarkable that the energy remains nearly constant throughout the calculations, which entail 234×10^6 updates of the smallest element in the mesh, or approximately 150 revolutions of the blade, despite the dynamically unstable initial transient. The energy behavior is equally good in the remaining numerical examples.

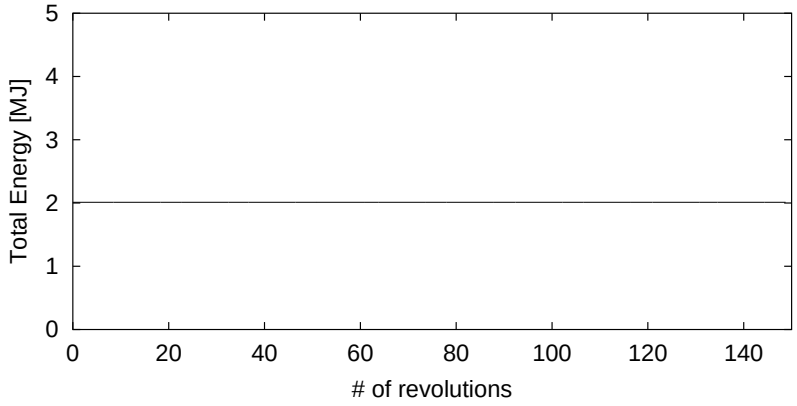


Figure 6.8: Evolution of the total energy in the blade as a function of the number of revolutions of the blade, for the third and softest case. Remarkably, the energy remains nearly constant even after the smallest element in the mesh has been updated more than 200 million times, at the end of the horizontal axis.

6.1.6 Complexity and convergence

The computational costs involved in the AVI algorithm can be separated into those associated with the element updates, and the overhead involved in the determination of which element to update next. It is readily verified, however, that the latter cost is generally much smaller than the former, as shown next.

Complexity and cost estimates. We proceed to estimate the speedup afforded by AVIs relative to explicit Newmark, or other similar explicit method, with a time step

$$\Delta t_{\min} = \min_K \Delta t_K.$$

For a fixed final time T the cost C_{AVI} of the AVI method is

$$C_{\text{AVI}} = \underbrace{\left(C_{\text{EU}} + C_{\text{PQ}}\right)}_{\text{cost per time step}} \cdot \underbrace{\sum_K \frac{T}{\Delta t_K}}_{\text{\# time steps}}, \tag{6.22}$$

where C_{EU} is the cost per element update and C_{PQ} is the cost of inserting and removing one element from the priority queue. The corresponding cost for explicit Newmark is

$$C_{\mathrm{NM}} = C_{\mathrm{EU}} \cdot \sum_K \frac{T}{\Delta t_{\min}}. \tag{6.23}$$

If we neglect the priority queue cost by setting $C_{\mathrm{PQ}} = 0$, then we can calculate the maximum AVI speedup to be

$$\lambda = \left.\frac{C_{\mathrm{NM}}}{C_{\mathrm{AVI}}}\right|_{C_{\mathrm{PQ}}=0} = \frac{\sum_K \frac{1}{\Delta t_{\min}}}{\sum_K \frac{1}{\Delta t_K}}. \tag{6.24}$$

We now consider the effect on the cost of the priority queue, which we assume to be implemented using a balanced binary tree. Under these conditions, the cost is $C_{\mathrm{PQ}} = \kappa \log_2(E)$, where E is the number of elements in the mesh and κ is the cost per level of the binary tree, which can be taken to be constant. Specifically, our benchmarks give $\kappa \approx 0.2\mu s$. By contrast, the cost C_{EU} of an element update depends sensitively on the type of element and the constitutive model, and can be substantial for complex constitutive models. Even for relatively simple nonlinear elasticity models, however, this cost is still greatly in excess of C_{PQ}. For example, for a Neohookean material and six noded triangles, our numerical benchmarks give $C_{\mathrm{EU}} \approx 10\mu s$. We can now estimate the size of the mesh for which the cost C_{PQ} begins to dominate C_{EU}. The ratio of costs is given by

$$\frac{C_{\mathrm{PQ}}}{C_{\mathrm{EU}}} = \frac{\kappa}{C_{\mathrm{EU}}} \log_2(E) = 0.02 \log_2(E),$$

and for C_{PQ} to be greater than C_{EU} would thus require $\log_2(E) > 50$, i.e., an inordinately large mesh containing in the order of 10^{15} elements. Taking the priority-queue overhead into consideration, true AVI speedup is

$$\frac{C_{\mathrm{NM}}}{C_{\mathrm{AVI}}} = \frac{\lambda}{1 + \frac{\kappa}{C_{\mathrm{EU}}} \log_2(E)}. \tag{6.25}$$

In the helicopter-blade calculations, the number of elements is $E = 2089$, the cost of an elemental update for a ten-node tetrahedral element is $C_{\mathrm{EU}} \approx 25\mu s$. and thus

$$\frac{C_{\mathrm{NM}}}{C_{\mathrm{AVI}}} = \frac{6.4}{1 + \frac{0.2}{25} \log_2(2089)} = 5.9,$$

which is close to the optimal speedup $\lambda = 6.4$. These estimates demonstrate the relatively modest impact of the priority queue on the overall performance of the method.

Probabilistic sorting for element updates. In §6.1.3 we considered solving the AVI equations (6.10) by maintaining a priority queue to track which element has the smallest next time in the entire mesh. This is an overly strong requirement for solving the equations, however, as we can update an element whenever all of the adjacent elements have larger next times. This methodology

also provides a basis for the construction of parallel implementations of AVI methods.

One simple scheme for doing this is to maintain the array of next times for each element. To perform a time step, we select an element at random and check whether its next time is less than that of all adjacent elements. If it is, then we update it and start another time step. If not, we move to the adjacent element with the smallest next time, and proceed to check its adjacent elements, repeating the test until we find an element with next time less than its neighbors. This procedure is clearly guaranteed to terminate, and it will always find an element which we can update.

The advantage of this method over the priority queue used earlier is that the cost will be constant in the number of elements in the mesh. To implement the method efficiently requires storing adjacency information in the mesh data structures, with which the cost of this algorithm will be very small compared with the other operations necessary to perform an elemental update.

To estimate the cost of this algorithm we note that in the worst case we may need to iterate over every element in the mesh before finding an element which we can update. This is extremely unlikely, however, and so it is appropriate to ask how many steps are necessary on average. We can readily calculate an upper bound for this by assuming an infinite mesh and that each element test is independent, which is overly pessimistic. For a mesh in which each element is adjacent to r other elements, the average number of steps before this algorithm finds an appropriate element is bounded by

$$\begin{aligned} N_{\text{step}} &= \sum_{n=1}^{\infty} n \left(1 - \left(\frac{1}{2}\right)^r\right) \left(1 - \left(\frac{1}{2}\right)^{r-1}\right)^{n-1} \left(\frac{1}{2}\right)^{r-1} \\ &= 2^{r-1} - \frac{1}{2}. \end{aligned} \tag{6.26}$$

For tetrahedral elements, such as those used in the numerical simulations in §6.1.5, each element has $r = 4$ adjacent elements and so the average number of steps necessary to find an element to update would be 7.5, irrespective of the size of the mesh. This algorithm would thus add a negligible cost to the AVI implementation and allow it to achieve almost the maximum speedup factor λ over explicit Newmark.

Cost/accuracy analysis. We quantitatively demonstrate that asynchronous methods can substantially lower the cost of a computation without degradation in accuracy by means of the following numerical test. Consider a two-dimensional plane strain, nonlinear elastic slab clamped on one edge and free at the other, deflecting under the action of uniform gravity. The material is Neohookean extended to the compressible range, eq. (6.20). The slab is initially undeformed and at rest. We consider an initial nonuniform mesh composed of 88 six-node quadratic triangles. A schematic picture of the geometry and the initial mesh are shown in Fig. 6.9.

The initial mesh is refined by uniform quadrisection several times over, resulting in meshes

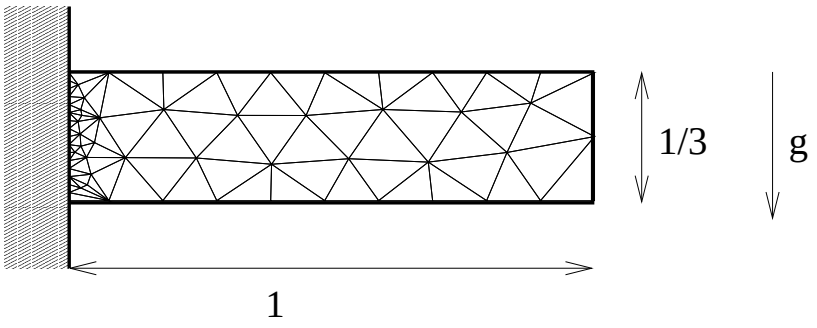


Figure 6.9: Schematic diagram of the geometry of the slab and the coarsest mesh for the cost/accuracy example

containing 88, 352, 1408, 5632 and 22528 elements. For each mesh, the solution is computed up to a prescribed final time using both AVIs and explicit Newmark. The explicit Newmark calculations are started as discussed in section §1.5.4. In the AVI calculations, the appropriate initial conditions at node a are

$$x_a^0 = \varphi_0(X_a)$$
$$v_a^{1/2} = \frac{1}{2m_a} \sum_{\{K|a\in K\}} I_{Ka}^0,$$

where $\varphi_0(X)$ is the initial deformation mapping and

$$I_{Ka}^0 = -\Delta t_K \frac{\partial V_K(x_K^0)}{\partial x_{Ka}^0}.$$

Note that this initialization procedure reduces to Newmark's for a uniform time step. An explicit Newmark solution computed with an exceedingly fine 90112-element mesh is presumed to be ostensibly converged, and used in lieu of the exact solution in order to compute errors.

The L^2-norm displacement and deformation-gradient errors are shown in Fig. 6.10 as a function of the computational cost. It is evident from these plots that the computational cost incurred in the computation of solutions of equal accuracy is is substantially less for the AVIs.

Convergence of AVIs. The slope of the plots in Fig. 6.10 is also directly related to the rate of convergence in Δt. To verify this, we note that the elemental time steps are proportional to the inner radius of the element, by virtue of the Courant condition. This implies that the rate of convergence in Δt is three-times the slope of the plots Fig. 6.10. This gives a displacement rate of convergence of roughly 1.9, and a deformation-gradient rate of convergence of 1.1, for both methods.

The convergence of AVIs may also be established analytically. To this end, define the ***maximum time step***

$$\Delta t_{\max} = \max_K \Delta t_K \tag{6.27}$$

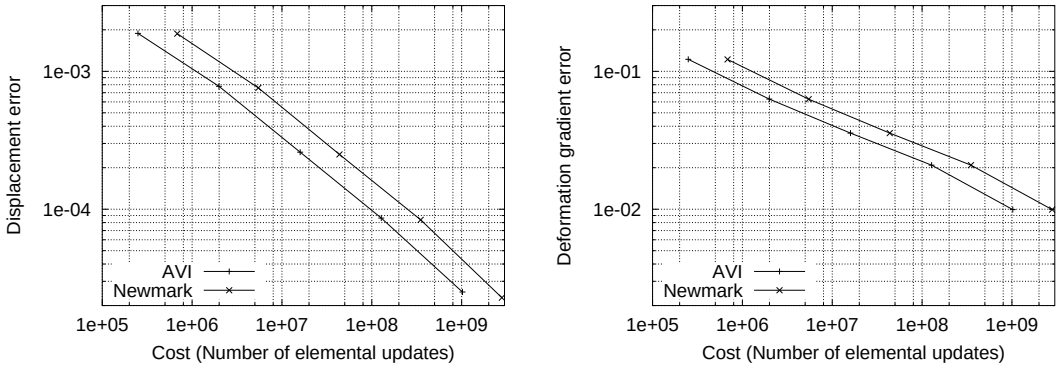


Figure 6.10: L^2 errors for the displacement (on the left) and the deformation gradient (on the right) as a function of the number of elemental updates for the slab problem. As is readily seen from the plots, AVIs are substantially cheaper than Newmark in computational cost for a desired error value.

and the ***maximum final time***

$$t_{\max} = \max_K N_K \Delta t_K. \tag{6.28}$$

With these definitions we can state the following result.

Theorem 6.1. *Consider a sequence of solutions obtained by the application of asychronous variational integrators to a fixed spatial discretization with maximum time step $\Delta t_{max} \to 0$ and maximum final time $t_{max} \to T$. Then the final discrete configuration converges to the exact configuration of the semidiscrete problem at time T.*

Proof. See §6.4 and Corollary 6.1. □

Resonance instabilities in asynchronous methods. It is known that multi-time step methods ocassionally exhibit resonance instabilities for some combinations of time steps (See, e.g., Izaguirre, Reich, and Skeel [1999], Bishop, Skeel, and Schulten [1997] and Biesiadecki and Skeel [1993] in the case of molecular dynamics, and Daniel [1997a] and Daniel [1997b] in the case of finite element methods in elastodynamics). We have not been able to detect any such instabilities in our simulations; however, a systematic study has yet to be performed.

6.2 Multisymplectic discretizations

Having investigated the variational multisymplectic structure of continuum mechanics and its associated conservation properties, we now turn to the general theory of constructing variational discretizations of such systems. The fundamental idea here is to discretize the variational structure, and then derive both the equations of motion (an integrator for the system) as well as conservation properties of the discrete system by using the same variational proofs as in the continuous case.

In this section we proceed in the same order as for the continuous case in §5. Namely, first we consider the discrete geometry of the problem, then define a discrete Lagrangian and a discrete variational principle and use these to derive first the Euler-Lagrange equations and then the conservation properties.

As we progress through this section we will develop both an abstract theory of variational discretizations, and simultaneously we will consider the example of AVI algorithms, concentrating on the geometry of the discrete problem.

6.2.1 Discrete Configuration Geometry

Discrete base spaces. A ***discrete base space configuration*** $\phi_{d,\mathcal{X}}$ consists of a set $\mathcal{X}_d$, called the ***nodal base space***, of points in $\mathcal{X}$ and a set $\mathcal{E}_d$ of subsets of $\mathcal{X}_d$, called the ***elemental base space***. Elements in $\mathcal{E}_d$ are regarded as encoding the connectivity between sets of nodes $X \in \mathcal{X}_d$, and we assume that we have a map $E \mapsto \mathcal{X}_E$ from elements $E \in \mathcal{E}_d$ to ***elemental subsets*** $\mathcal{X}_E$ of $\mathcal{X}$. We write $\mathcal{X}_{\mathcal{E}_d} = \cup_{E \in \mathcal{E}_d} \mathcal{X}_E$ for the subset of $\mathcal{X}$ covered by the elemental subsets. Given a node $X \in \mathcal{X}_d$ we denote by $\mathcal{E}_d(X)$ the set of elements containing that node, so that $\mathcal{E}_d(X) = \{E \in \mathcal{E}_d \mid X \in E\}$.

Note that the elements and nodes referred to here are *space-time* elements and nodes. That is, each elemental subset is a subset of space and time, while each node specifies both a spatial position and a particular time. This is in contrast to the normal usage in finite elements, where the terms element and node refer solely to spatial objects. We also do not necessarily consider a set of basis functions over the elements, as we may wish to use different discretization schemes in some components, such as finite differences for time derivatives.

For discussing boundary conditions and equations it is necessary to specify the boundary and interior of the nodal base space. These are, respectively,

$$\begin{aligned} \partial \mathcal{X}_d &= \{X \in \mathcal{X}_d \mid X \in \partial \mathcal{X}_{\mathcal{E}_d}\} \\ \text{int}(\mathcal{X}_d) &= \mathcal{X}_d \setminus \partial \mathcal{X}_d. \end{aligned}$$

We denote by $\mathcal{C}_d(\mathcal{X})$ the space of all allowed discrete base space configurations $\phi_{d,\mathcal{X}}$, which we will take to all have the same number of nodes and elements. Note that we will generally not be allowing arbitrary nodal base spaces, but will rather impose some restrictions on the configurations under consideration.

AVI base spaces. In the particular case of AVI methods, we assume a fixed reference mesh $\mathcal{T}$, and so the space of discrete base space configurations is parameterized by the set of elemental times t_K^j. We assume that we have a fixed spatial discretization, as in §6.1.1. For given elemental times

t_K^j and induced nodal times the corresponding discrete nodal and elemental base spaces are

$$\mathcal{X}_d = \{X_a^i = (t_a^i, X_a) \mid a \in \mathcal{T},\ 1 \le i \le N_a\}$$
$$\mathcal{E}_d = \left\{E_K^j = \{X_a^i \mid a \in K,\ t_a^i \in \Theta^{K,j}\} \;\middle|\; K \in \mathcal{T},\ 1 \le j < N_K\right\}.$$

The map from an element E to a subset $\mathcal{X}_E$ for AVI methods is given by $\mathcal{X}_{E_K^j} = [t_K^j, t_K^{j+1}] \times K$.

Discrete configuration bundles. Having defined discrete base space configurations, we now turn to constructing discrete representations of the configuration bundle $\pi_{\mathcal{X}Y} : Y \to \mathcal{X}$. For a given $\phi_{d,\mathcal{X}}$, we define the ***discrete configuration bundle*** Y_d to be the fiber bundle over $\mathcal{X}_d$ with the fiber over $X \in \mathcal{X}_d$ being simply the configuration bundle fiber Y_X itself.

A ***discrete configuration*** ϕ_d now consists of a discrete base space configuration $\phi_{d,\mathcal{X}}$ and a section of Y_d. Such a section can also be regarded as a map $\mathcal{X}_d \to Y$ covering the identity. A discrete configuration ϕ_d thus specifies a set of nodes $\mathcal{X}_d$, a set of elements $\mathcal{E}_d$, and a fiber value denoted x_X at each node $X \in \mathcal{X}_d$.

AVI configuration bundle. For AVI methods, we have seen above that the discrete nodal and elemental spaces which make up the discrete base space configuration $\phi_{d,\mathcal{X}}$ are specified by the times t_K^j. An AVI configuration ϕ_d thus consists of these sets, together with the fiber positions x_a^i for each node $X_a^i \in \mathcal{X}_d$.

We denote the set of all allowable discrete configurations by $\mathcal{C}_d(Y)$. This is the space of allowable discrete base space configuration $\mathcal{C}_d(\mathcal{X})$ together with the product of as many fibers Y_X as there are nodes.

Discrete jet bundle. One of the fundamental foundations of the discrete approach is to replace continuous derivative information with a finite collection of samples of a function. To formulate this more precisely, for a given discrete base space configuration $\phi_{d,\mathcal{X}}$ we define the ***discrete jet bundle*** to be the fiber bundle J^1Y_d over $\mathcal{E}_d$ where the fiber over $E \in \mathcal{E}_d$ is the product of the fibers over each node in E. That is,

$$(J^1Y_d)_E = \prod_{X \in E} Y_X.$$

Each point in the discrete jet bundle thus stores the value of the configuration at all nodes of the given element.

Given a discrete configuration ϕ_d we define the ***discrete jet extension*** $j^1\phi_d$ to be the section of J^1Y_d specified by $j^1\phi_d(E) = (E, \{x_X \mid X \in E\})$, which is simply the configuration evaluated at all nodes within a single element. This is enough information to form discrete approximations to the derivative.

AVI discrete jet bundle. For AVI methods, we have seen that the a discrete base space configuration consists of nodes $X_a^i = (t_a^i, X_a)$ and elements E_K^j. The discrete configuration bundle then consists of all possible spatial positions for each material node X_a at each time t_a^i. The corresponding discrete jet bundle therefore consists of elements E_K^j, specifying a material element K and times t_K^j, t_K^{j+1}, together with the set of possible spatial positions for each node $X_a \in K$ at each time $t_a^i \in \Theta_j^K$. The discrete jet extension of a discrete configuration ϕ_d is thus given by

$$\begin{aligned} j^1\phi_d(E_K^j) &= \Big(E_K^j, \{x_a^i \mid X_a^i \in E_K^j\}\Big) \\ &= \Big(E_K^j, \{x_a^i \mid a \in K,\ t_a^i \in \Theta_j^K\}\Big). \end{aligned}$$

Discrete Lagrangian. To complete the specification of the discrete system, we must now provide a discrete equivalent of the Lagrangian function, namely a ***discrete Lagrangian*** $L_d : J^1Y_d \to \mathbb{R}$. This should not approximate the continuous Lagrangian, however, but rather should be thought of as an approximation to the continuous action integral over a single element. That is,

$$L_d\big(E, \{x_X \mid X \in E\}\big) \approx \int_{\mathcal{X}_E} L(j^1\varphi)\ \mathbf{d}^{n+1}X,$$

where φ is an exact solution of the Euler-Lagrange equations for L over the elemental subset $\mathcal{X}_E$ which is approximated by the fiber values x_X at the nodes $X \in E$. We will frequently use the shorthand notation $L_d(E) = L_d(E, \{x_X \mid X \in E\})$ for the arguments of the discrete Lagrangian.

Example of AVI discrete Lagrangian. We have seen that a single point in the AVI discrete jet bundle consists of an element E_K^j, consisting of the nodes $X_a^i = (t_a^i, X_a)$, together with the spatial positions x_a^i corresponding to each node. The nodal times include the elemental times t_K^j and t_K^{j+1}, so a discrete jet bundle point is precisely the quantities on which the AVI discrete Lagrangian (6.8) is defined. This clearly approximates the action over the elemental subset $\mathcal{X}_{E_K^j} = [t_K^j, t_K^{j+1}] \times K$.

6.2.2 Discrete Variations and Dynamics

Discrete variations. We first consider horizontal variations. The space of variations of a discrete base space configuration $\phi_{d,\mathcal{X}}$ is the tangent space $T_{\phi_{d,\mathcal{X}}}\mathcal{C}_d(\mathcal{X})$, with each variation being a map $\delta\phi_{d,\mathcal{X}} : \mathcal{X}_d \to T\mathcal{X}$ covering the identity. Here we will assume that the elemental base space does not alter its connectivity, and thus moves along with the nodes. It will be important below to distinguish between boundary variations and interior variations. We thus assume that we can write the tangent space as a direct sum

$$T_{\phi_{d,\mathcal{X}}}\mathcal{C}_d(\mathcal{X}) = T^i_{\phi_{d,\mathcal{X}}}\mathcal{C}_d(\mathcal{X}) \oplus T^\partial_{\phi_{d,\mathcal{X}}}\mathcal{C}_d(\mathcal{X})$$

of interior and boundary components, respectively. We write $\pi^i_{\mathcal{X}}$ and $\pi^\partial_{\mathcal{X}}$ for the associated projections, and for a given variation $\delta\phi_{d,\mathcal{X}}$ we denote the two components by $\delta_i\phi_{d,\mathcal{X}} = \pi^i_{\mathcal{X}} \cdot \delta\phi_{d,\mathcal{X}}$ and $\delta_\partial\phi_{d,\mathcal{X}} = \pi^\partial_{\mathcal{X}} \cdot \delta\phi_{d,\mathcal{X}}$.

Now we define full (vertical and horizontal) variations. Similarly to the above, the space of variations of a discrete configuration ϕ_d is the tangent space $T_{\phi_d}\mathcal{C}_d(Y)$ consisting of variations $\delta\phi_d : \mathcal{X}_d \to TY$ covering the section of Y_d. This decomposes naturally into a horizontal base space component and a vertical component, according to

$$\begin{aligned} T_{\phi_d}\mathcal{C}_d(Y) &= T_{\phi_{d,\mathcal{X}}}\mathcal{C}_d(\mathcal{X}) \oplus T^V_{\phi_d}\mathcal{C}_d(Y) \\ T^V_{\phi_d}\mathcal{C}_d(Y) &= \bigoplus_{X \in \mathcal{X}_d} T_{x_X} Y_X. \end{aligned}$$

The vertical component of a variation can thus be written as a sum of variations of each fiber variable, which we denote by $\delta x_X \in T_{x_X}Y_X$ for each $X \in \mathcal{X}_d$. We will abuse the notation and also write δx_X and $\delta_{i,\partial}\phi_{d,\mathcal{X}}$ for the relevant projections in $T_{\phi_d}\mathcal{C}_d(Y)$. This gives a full decomposition of a variation into the vertical interior, vertical boundary, horizontal interior and horizontal boundary components as

$$\delta\phi_d = \sum_{X \in \mathrm{int}(\mathcal{X}_d)} \delta x_X + \sum_{X \in \partial\mathcal{X}_d} \delta x_X + \delta_i\phi_{d,\mathcal{X}} + \delta_\partial\phi_{d,\mathcal{X}}. \tag{6.29}$$

Boundary and interior variations differ in a key property. Interior variations are zero on all $X \in \partial\mathcal{X}_{\mathcal{E}_d}$, whereas boundary variations have nonzero components on the boundary.

Variations of AVI configurations. Given an AVI configuration ϕ_d and a variation $\delta\phi_d$ of it, we can decompose it as above into horizontal components and vertical per-fiber components. We can also, however, take advantage of the special structure of the AVI configuration bundles to further decompose the horizontal components.

An AVI base space configuration $\phi_{d,\mathcal{X}}$ is specified by the elemental times t^j_K, so variations in the configuration are induced by variations in the times. We denote by $\delta^j_K\phi_{d,\mathcal{X}}$ the configuration variation induced by δt^j_K, and we take the boundary variations to be those associated with times t^1_K and $t^{N_K}_K$. This provides a decomposition of any variation of an AVI configuration into

$$\delta\phi_d = \sum_{X^i_a \in \mathrm{int}(\mathcal{X}_d)} \delta x^i_a + \sum_{X^i_a \in \partial\mathcal{X}_d} \delta x^i_a + \sum_{K \in \mathcal{T}} \sum_{1<j<N_K} \delta^j_K\phi_{d,\mathcal{X}} + \sum_{K \in \mathcal{T}} (\delta^1_K\phi_{d,\mathcal{X}} + \delta^{N_K}_K\phi_{d,\mathcal{X}}). \tag{6.30}$$

Discrete Euler-Lagrange equations. To formulate a discrete variational principle, we begin by defining the ***discrete action sum*** $S_d : \mathcal{C}_d(Y) \to \mathbb{R}$ to be

$$S_d(\phi_d) = \sum_{E \in \mathcal{E}_d} L_d\big((j^1\phi_d)(E)\big). \tag{6.31}$$

We can now formulate the ***discrete Hamilton's principle***, which states that we must seek critical points of the discrete action function. That is, we say that $\bar{\phi}$ is a ***discrete solution*** if

$$\mathbf{d}S_d(\phi_d) \cdot \delta\phi_d = 0$$

for all variations $\delta\phi_d$ with zero boundary components. We will write D_V and D_H for the derivatives with respect to vertical and horizontal components, respectively, so that using the above decomposition (6.29) of variations gives

$$\begin{aligned}
\mathbf{d}S_d(\phi_d) \cdot \delta\phi_d &= \sum_{E \in \mathcal{E}_d} \sum_{X \in E} \frac{\partial L_d(E)}{\partial x_X} \delta x_X + D_H S_d(\phi_d) \cdot \delta\phi_{d,\mathcal{X}} \\
&= \sum_{X \in \mathrm{int}(\mathcal{X}_d)} \left(\sum_{E \in \mathcal{E}_d(X)} \frac{\partial L_d(E)}{\partial x_X} \right) \cdot \delta x_X + D_H S_d(\phi_d) \cdot \delta_i \phi_{d,\mathcal{X}} \\
&\quad + \sum_{X \in \partial\mathcal{X}_d} \left(\sum_{E \in \mathcal{E}_d(X)} \frac{\partial L_d(E)}{\partial x_X} \right) \cdot \delta x_X + D_H S_d(\phi_d) \cdot \delta_\partial \phi_{d,\mathcal{X}}.
\end{aligned} \tag{6.32}$$

The requirement that this expression be zero for all nonzero interior variations implies that the first two terms must be zero. The first of these, arising from vertical variations, is termed the ***discrete Euler-Lagrange equations***:

$$\sum_{E \in \mathcal{E}_d(X)} \frac{\partial L_d(E)}{\partial x_X} = 0 \tag{6.33}$$

for all $X \in \mathrm{int}(\mathcal{X}_d)$. This is a finite set of equations which relate the configuration variables making up ϕ_d. We will investigate the second term in (6.32) below.

Observe that we obtain one discrete Euler-Lagrange equation per fiber configuration variable x_X associated to an internal node $X \in \mathrm{int}(\mathcal{X}_d)$. If we thus regard both the base space configuration $\phi_{d,\mathcal{X}}$ and the fiber variables variables x_X for $X \in \partial\mathcal{X}_d$ as fixed, then the discrete Euler-Lagrange equations are sufficient, at least in terms of an equation count, to uniquely solve for a discrete configuration ϕ_d.

Equations for AVI methods. Requiring that the discrete AVI action is stationary with respect variations in the configuration variables x_a^i for internal nodes X_a^i gives the equations

$$\sum_{E_K^j \in \mathcal{E}_d(X_a^i)} \frac{\partial L_d(E_K^j)}{\partial x_a^i} = 0.$$

For the discrete Lagrangian (6.8) we have already calculated this explicitly in §6.1.4.

Boundary conditions. As in the continuous problem, we consider zeroth and first-order boundary conditions of the form

$$x_X = x_0(X) \quad \text{for } X \in \partial_0 \mathcal{X}_d \tag{6.34a}$$

$$\sum_{E \in \mathcal{E}_d(X)} \frac{\partial L_d(E)}{\partial x^a} = \tau_a(X) \quad \text{for } X \in \partial_1 \mathcal{X}_d, \tag{6.34b}$$

where $\partial_0 \mathcal{X}_d$ and $\partial_1 \mathcal{X}_d$ are subsets of the discrete nodal space boundary $\partial \mathcal{X}_d$, and x_0 and τ are given functions. We do not require that $\partial_0 \mathcal{X}_d$ and $\partial_1 \mathcal{X}_d$ be disjoint, nor that they cover $\partial \mathcal{X}_d$. Note that this τ will typically only be an approximation to the τ in the continuous case.

We impose the boundary conditions by modifying the discrete Hamilton's principle to seek discrete configurations ϕ_d satisfying (6.34a) for which

$$\mathbf{d}S_d(\phi_d) \cdot \delta\phi_d = \sum_{X \in \partial_1 \mathcal{X}} \tau(X) \cdot \delta x_X \tag{6.35}$$

for all variations $\delta\phi_d$ of ϕ_d which are zero on the set $\partial \mathcal{X}_d \setminus \partial_1 \mathcal{X}_d$. This is exactly analogous to the way we imposed boundary conditions for the continuous problem in §5.1.2.

AVI boundary conditions. In applications of the AVI method we are generally concerned with initial boundary value problems (IBVP), for which the boundary conditions are given as

$$x_a^1 = (x_0)_a^1 \qquad \text{for all nodes } a \in \mathcal{T} \tag{6.36a}$$

$$\sum_{K \in \mathcal{T}_a} \frac{\partial L_d(E_K^1)}{\partial x_a^{i_a(K,1)}} = -(p_0)_a \quad \text{for all nodes } a \in \mathcal{T} \tag{6.36b}$$

$$x_a^i = (x_0)_a^i \qquad \text{for all } i = 1, \ldots, N_a,\ X_a \in \partial_d \mathcal{B} \tag{6.36c}$$

$$\sum_{K \in \mathcal{T}_a} \sum_{\substack{j \\ t_a^i \in \Theta^{K,j}}} \frac{\partial L_d(E_K^j)}{\partial x_a^i} = \tau_a^i \qquad \text{for all } i = 1, \ldots, N_a,\ X_a \in \partial_\tau \mathcal{B}. \tag{6.36d}$$

In the context of solid mechanics, the first two of these are termed the initial conditions and the final two are termed boundary conditions. The initial conditions are both zeroth- and first-order boundary conditions, and so we have the space-time boundaries

$$\partial_0 \mathcal{X}_d = \left\{ X_a^{i_a(1,K)} \mid a \in \mathcal{T},\ K \in \mathcal{T}_a \right\} \cup \left\{ X_a^i \mid a \in \partial_d \mathcal{T},\ 1 \le i \le N_a \right\}$$

$$\partial_1 \mathcal{X}_d = \left\{ X_a^{i_a(1,K)} \mid a \in \mathcal{T},\ K \in \mathcal{T}_a \right\} \cup \left\{ X_a^i \mid a \in \partial_\tau \mathcal{T},\ 1 \le i \le N_a \right\}.$$

6.2.3 Horizontal Variations

In continuous multisymplectic mechanics, we have seen that horizontal variations give equations which are functionally dependent on the Euler-Lagrange equations derived from vertical variations, and so they may be considered as conservation laws of the system.

This is not the case once the system has been discretized. Indeed, requiring stationarity with respect to horizontal variations for the discrete system gives new equations which can be used to solve for the discrete base space configuration, and thus for the space-time mesh. Both space and time adaptivity could eventually be driven by this set of discrete equations.

More precisely, from the discrete Hamilton's principle and equation (6.32) for the action variations, we see that interior horizontal variations give the equations

$$D_H S_d(\phi_d) \cdot \delta_i \phi_{d,\mathcal{X}} = 0 \tag{6.37}$$

for all $\delta_i \phi_{d,\mathcal{X}} \in T^i_{\phi_{d,\mathcal{X}}} \mathcal{C}_d(\mathcal{X})$. As there is one equation arising from each interior horizontal variation, these equations are sufficient to solve for $\phi_{d,\mathcal{X}}$ given appropriate boundary conditions.

It is important to be clear that equation (6.37) is not simply a conservation law for a system satisfying (6.33), but is an independent set of equations. Nonetheless, this equation can also be regarded as enforcing the conservation of discrete quantities corresponding to continuous horizontal conserved quantities.

AVI methods and energy conservation. For AVI methods we have seen that the discrete base space configurations $\phi_{d,\mathcal{X}}$ are parameterized by the space of elemental times t^j_K, and that these also parameterize the space of horizontal variations. Requiring that the action be stationary with respect to the variation $\delta^j_K \phi_{d,\mathcal{X}}$ associated with each interior time t^j_K for $1 < j < N_K$ gives a local energy conservation equation. Summing over all elements $K \in \mathcal{T}$ then gives a discrete global energy conservation equation as a consequence, which is the discrete analogue of equation (5.14). We will also see below how this may be viewed as a consequence of the discrete Noether's theorem.

In the AVI method of §6.1 we have taken the set of allowed discrete base space configurations to be those with space-time nodes of the form $X^i_a = (t^i_a, X_a)$ for fixed material nodes X_a. One could think of taking a larger class of base space configurations, where the spatial coordinates of each X^i_a were allowed to vary independently. The nodal times would still be induced by the elemental times t^j_K, so the set of space-time meshes would be parameterized by the t^j_K and positions $X^i_a \in \mathcal{B}$ for each node a and time t^i_a. Requiring stationarity of the action with respect to the times would still give discrete energy conservation, and we could additionally require stationarity with respect to the horizontal spatial nodal variations. This would give discrete configuration forces, in analogy to §5.1.3.

We shall see in §6.3.1 and §6.3.2 that the multisymplectic nature of the discrete algorithm does *not* depend on requiring stationarity with respect to horizontal variations. A similar statement holds for the discrete Noether's theorem.

6.3 Discrete conservation laws

We will now see how the variational derivations of the conservation laws for continuous multisymplectic systems carry over directly to variational multisymplectic discretizations.

Discrete space of solutions. Recall that $\mathcal{C}_d(Y)$ denotes the space of discrete configurations ϕ_d. By $\mathcal{C}_{L_d}(Y)$ we denote the ***discrete space of solutions***, which is all configurations ϕ_d which satisfy the discrete Euler-Lagrange equations for some boundary conditions. Tangent vectors $V_d \in T_{\phi_d}\mathcal{C}_{L_d}(Y)$ are called ***discrete first variations*** and are derivatives of a curve of solutions. We write the decomposition of V_d according to (6.29) as

$$V_d = \sum_{X \in \text{int}(\mathcal{X}_d)} V_{d,x_X} + \sum_{X \in \partial\mathcal{X}_d} V_{d,x_X} + V^i_{d,\mathcal{X}} + V^\partial_{d,\mathcal{X}}$$

into the interior vertical, boundary vertical, interior horizontal, and boundary horizontal components, respectively. We will also use the notation $V^i_{d,V}$ and $V^\partial_{d,V}$ to denote the entire interior vertical and boundary vertical terms above. Given a discrete variation V_d we can construct its jet extension j^1V_d, which takes E to the set of variations $V_d(X)$ for each $X \in E$.

It is often useful to consider different spaces of solutions corresponding to the requirement of action stationarity with respect to different classes of variations. For example, we could consider the space of solutions for the AVI algorithm with only the discrete Euler-Lagrange equations arising from vertical variations satisfied, or we could consider the space of solutions to also have the requirement of stationarity with respect to horizontal variations. In either case we will have a discrete multisymplectic form formula and discrete Noether's theorem, but the exact expression of each will differ for the different solution spaces. Here we will write the expressions in the general case of full vertical and horizontal variations, so that the expressions for vertical only solutions can be obtained by dropping the horizontal terms. While this provides the most generality, we should remember that the numerical examples from §6.1.5 were performed using the AVI algorithm without considering horizontal variations.

6.3.1 Discrete Multisymplectic Forms

One of the powerful features of variational multisymplectic discretizations is that there is a unique discrete multisymplectic structure defined by a given discretization. This appears as the boundary

term in free action variations, just as in the continuous case.

Equations (6.32) and (6.37) show that restricting to the space of solutions eliminates the interior terms, and so we can write

$$\mathbf{d}S_d(\phi_d)\cdot V_d = \sum_{X\in\partial\mathcal{X}_d}\sum_{E\in\mathcal{E}_d(X)}\Theta_{L_d}^{E,X}(j^1\phi_d(E))\cdot j^1V_d + D_HS_d(\phi_d)\cdot V_{d,\mathcal{X}}^{\partial} \tag{6.38}$$

for all solutions ϕ_d and first variations V_d. Here $\Theta_{L_d}^{E,X}$ are the ***discrete Cartan forms*** defined by

$$\Theta_{L_d}^{E,X} = \frac{\partial L_d(E)}{\partial x_X}\,\mathbf{d}x_X.$$

As in the continuous case, we now define the ***discrete multisymplectic Lagrangian forms*** $\Omega_{L_d}^{E,X}$ to be the exterior derivatives of the corresponding discrete Cartan forms with respect to vertical variables

$$\Omega_{L_d}^{E,X} = -\mathbf{d}_V\Theta_{L_d}^{E,X}.$$

Calculating this explicitly gives

$$\Omega_{L_d}^{E,X} = -\sum_{X'\in E\setminus X}\frac{\partial^2 L_d(E)}{\partial x_{X'}\partial x_X}\mathbf{d}x_{X'}\wedge\mathbf{d}x_X.$$

6.3.2 Discrete Multisymplectic Form Formula

Taking a second exterior derivative of the action derivative expression (6.38) and using $\mathbf{d}^2=0$ now immediately gives the ***discrete multisymplectic form formula***

$$\sum_{X\in\partial\mathcal{X}_d}\sum_{E\in\mathcal{E}_d(X)}\Omega_{L_d}^{E,X}(j^1\phi_d(E))\cdot(j^1V_d,j^1W_d) + D_VD_HS_d(\phi_d)\cdot V_{d,\mathcal{X}}^{\partial}\cdot W_d^{\partial} + D_HD_VS_d(\phi_d)\cdot V_{d,V}^{\partial}\cdot W_{d,\mathcal{X}}^{\partial} + D_HD_HS_d(\phi_d)\cdot V_{d,\mathcal{X}}^{\partial}\cdot W_{d,\mathcal{X}}^{\partial} = 0$$

for all discrete first variations V_d and W_d. This is a discretization of the expression (5.17) of the continuous multisymplectic form formula.

If we repeat this calculation for a single element rather than the entire configuration, we obtain the ***discrete local multisymplectic form formula***

$$\sum_{X\in E}\Omega_{L_d}^{E,X}(j^1\phi_d(E))\cdot(j^1V_d,j^1W_d) + D_VD_HL_d(E)\cdot V_{d,\mathcal{X}}^{\partial}\cdot W_d^{\partial} + D_HD_VL_d(E)\cdot V_{d,V}^{\partial}\cdot W_{d,\mathcal{X}}^{\partial} + D_HD_HL_d(E)\cdot V_{d,\mathcal{X}}^{\partial}\cdot W_{d,\mathcal{X}}^{\partial} = 0$$

for any element E, and all discrete variations V_d and W_d (not necessarily first variations). This

expression is a discretization of the divergence form (5.18) of the continuous multisymplectic form formula, and summing over all elements and using the discrete Euler-Lagrange equations will give the above global form.

If we are considering only vertical variations, then the global and local discrete multisymplectic form formulas simplify to give just

$$0 = \sum_{X \in \partial \mathcal{X}_d} \sum_{E \in \mathcal{E}_d(X)} \Omega_{L_d}^{E,X}(j^1\phi_d(E)) \cdot (j^1 V_d, j^1 W_d)$$
$$0 = \sum_{X \in E} \Omega_{L_d}^{E,X}(j^1\phi_d(E)) \cdot (j^1 V_d, j^1 W_d)$$

for solutions ϕ_d and first variations V_d and W_d.

6.3.3 Discrete Reciprocity and Time Symplecticity

In the continuous case we have seen that the multisymplectic form formula is a generalization of the notions of reciprocity for static problems and time-symplecticity for dynamic problems into one single space-time statement. In the discrete case this is also true, and so by restricting the above statements to particular classes of variations we can recover exact discrete reciprocity and exact symplecticity in time for variational discretizations.

Discrete reciprocity. Consider now a discrete problem with only vertical variations. A linearized solution W_d about ϕ_d of the discrete system (6.35) for the incremental body force B_d^W and incremental traction τ_d^W satisfies

$$D_V(D_V S_d(\phi_d) \cdot V_d) \cdot W_d = \sum_{X \in \mathcal{X}_d} B_d^W(X) \cdot V_d(X) + \sum_{X \in \partial_\tau \mathcal{X}_d} \tau_d^W(X) \cdot V_d(X)$$

for all variations V_d which are zero on the displacement boundary. The identity $D_V(D_V S_d(\phi_d) \cdot V_d) \cdot W_d = D_V(D_V S_d(\phi_d) \cdot W_d) \cdot V_d$ holds for discrete as well as continuous systems, and so we immediately obtain the relation

$$\sum_{X \in \mathcal{X}_d} B_d^W(X) \cdot V_d(X) + \sum_{X \in \partial_\tau \mathcal{X}_d} \tau_d^W(X) \cdot V_d(X) = \sum_{X \in \mathcal{X}_d} B_d^V(X) \cdot W_d(X) + \sum_{X \in \partial_\tau \mathcal{X}_d} \tau_d^V(X) \cdot W_d(X).$$

This is exactly a discrete reciprocity law, as can be seen by comparing it to the continuous version in §5.2.3.

The interpretation is the same as in the continuous case, with applied forces B_d^W and τ_d^W producing the linearized response W, and similarly for V. Then measuring V in the direction of the

forces B_d^W, τ_d^W gives precisely the same result as measuring the response W in the direction B_d^V, τ_d^V.

This is equivalent to symmetry of the stiffness matrix, which, as is well known, results automatically from a variational discretization.

Discrete time symplecticity. We now turn to considering an initial boundary value problem such as that specified by (6.35) for the conditions (6.36) with $\tau = 0$, and we restrict ourselves to vertical variations. Consider a smooth curve of initial conditions $(x_i^\epsilon, p_i^\epsilon)$ which is (x_i, p_i) at $\epsilon = 0$, and let ϕ_d^ϵ be the corresponding solutions for all time. Given a variation in the initial conditions of the form

$$(\delta x_i, \delta p_i) = \left.\frac{\partial}{\partial \epsilon}\right|_{\epsilon=0} (x_i^\epsilon, p_i^\epsilon),$$

we induce a variation of the solution by

$$V_d = \left.\frac{\partial}{\partial \epsilon}\right|_{\epsilon=0} \phi_d^\epsilon.$$

We also consider a discrete flow map F_{L_d} which maps from initial conditions (x_i, p_i) to final conditions (x_f, p_f) of the system. The variation $(\delta x_f, \delta p_f)$ corresponding to $(\delta x_i, \delta p_i)$ then satisfies

$$(\delta x_f, \delta p_f) = TF_{L_d} \cdot (\delta x_i, \delta p_i).$$

Now consider a second variation $(\delta' x_i, \delta' p_i)$ which induces V_d' and $(\delta' x_f, \delta' p_f)$. We assume a decomposition of the boundary $\partial \mathcal{X}_d = \partial_i \mathcal{X}_d \cup \partial_f \mathcal{X}_d \cup \partial_d \mathcal{X}_d \cup \partial_\tau \mathcal{X}_d$ into the initial, final, spatial displacement boundary and spatial traction boundary components, respectively. These sets are all disjoint, and together they cover $\partial \mathcal{X}_d$. The variations V_d and W_d are zero on $\partial_d \mathcal{X}_d$ and on $\partial_\tau \mathcal{X}_d$ we have $\tau = 0$, so the multisymplectic form formula becomes

$$\sum_{X \in \partial_i \mathcal{X}_d} \sum_{E \in \mathcal{E}_d(X)} \Omega_{L_d}^{E,X}(j^1\phi_d(E)) \cdot (j^1 V_d, j^1 W_d) + \sum_{X \in \partial_f \mathcal{X}_d} \sum_{E \in \mathcal{E}_d(X)} \Omega_{L_d}^{E,X}(j^1\phi_d(E)) \cdot (j^1 V_d, j^1 W_d) = 0.$$

We now define the discrete field theoretic two-forms

$$\Omega_{L_d}^i(\delta x_i, \delta p_i) = -\sum_{X \in \partial_i \mathcal{X}_d} \sum_{E \in \mathcal{E}_d(X)} \Omega_{L_d}^{E,X}(j^1\phi_d(E)) \cdot (j^1 V_d, j^1 W_d)$$

$$\Omega_{L_d}^f(\delta x_f, \delta p_f) = \sum_{X \in \partial_f \mathcal{X}_d} \sum_{E \in \mathcal{E}_d(X)} \Omega_{L_d}^{E,X}(j^1\phi_d(E)) \cdot (j^1 V_d, j^1 W_d)$$

and so using the fact that the initial and final variations are related by TF_{L_d} we have

$$\Omega^i_{L_d} = (F_{L_d})^* \Omega^f_{L_d}, \tag{6.39}$$

which is exactly a discretization of the continuous equivalent (5.19).

Note that we could also consider both vertical and time-horizontal variations in the derivation of the above relationship. This would then give a discrete analogue of extended time-symplecticity, namely the preservation of the two-form $\mathbf{d}\left(\frac{\partial L}{\partial \dot{q}^a}\right) \wedge \mathbf{d}p_a + \mathbf{d}E_L \wedge \mathbf{d}t$ (see Marsden and West [2001] for the details of this in the case of ODEs).

For AVIs equation (6.39) encodes a generalized type of time symplecticity. Note that this does not mean that we can use standard backward error methods for analyzing AVIs, as we do not have a single symplectic form on a space with an iterated symplectic map. Nonetheless, we conjecture that it is the geometric property (6.39) which is responsible for the excellent energy behavior observed numerically for AVI methods.

6.3.4 Discrete Noether's Theorem

We now develop a discrete Noether's theorem associated to vertical variations. Take a group action $\Phi : G \times Y \to Y$, as in §5.2.5, which acts by diffeomorphisms $g : Y \to Y$ covering diffeomorphisms $g_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}$. The corresponding infinitesimal generators are ξ_Y and $\xi_{\mathcal{X}}$, as defined previously.

We may also consider G as acting on the discrete configuration bundle by pointwise action on Y_d, so the infinitesimal generators $\xi_{\mathcal{X}_d}$ and ξ_{Y_d} are pointwise equal to $\xi_{\mathcal{X}}$ and ξ_Y. Given a discrete base space configuration $\phi_{d,\mathcal{X}}$, we define the action of G on $\mathcal{C}_d(\mathcal{X})$ to be pointwise action on the nodal positions $\mathcal{X}_d$, and we assume that the elemental topology specified by $\mathcal{E}_d$ is left invariant. We similarly define the action of G on $\mathcal{C}_d(Y)$ by the action on $\mathcal{C}_d(\mathcal{X})$ together with pointwise action on the fibers. Here we implicitly assume that the action of G is such that it preserves the space $\mathcal{C}_d(\mathcal{X})$. That is, for any allowed base space configuration $\phi_{d,\mathcal{X}} \in \mathcal{C}_d(\mathcal{X})$, the transformed base space configuration $g \cdot \phi_{d,\mathcal{X}}$ is also an allowed configuration, and thus $g \cdot \phi_{d,\mathcal{X}} \in \mathcal{C}_d(\mathcal{X})$.

The action of G on Y_d can be prolonged to the discrete jet bundle $J^1 Y_d$ by pointwise action on each component, which means that the corresponding infinitesimal generator $\xi_{J^1 Y_d} : J^1 Y_d \to T(J^1 Y_d)$ is a vector

$$\xi_{J^1 Y_d}\Big(E, \{x_X \mid X \in E\}\Big) = \Big(E, \{x_X \mid X \in E\}, \xi_X(E), \{\xi_Y(x_X) \mid X \in E\}\Big)$$

consisting of pointwise evaluations of ξ_Y. We will denote the vertical components of this by

$$\xi^V_{J^1 Y_d}\Big(E, \{x_X \mid X \in E\}\Big) = \Big(E, \{x_X \mid X \in E\}, 0, \{\xi_Y(x_X) \mid X \in E\}\Big).$$

A group action is said to be a ***symmetry*** of the discrete Lagrangian L_d if

$$L_d(E, \{x_X \mid X \in E\}) = L_d\Big(g \cdot (E, \{x_X \mid X \in E\})\Big)$$

for all point in J^1Y_d and all $g \in G$, and in such a case the discrete Lagrangian is said to be ***equivariant***. This implies that the discrete Lagrangian is ***infinitesimally equivariant***, which is the requirement

$$\mathbf{d}L_d \cdot \xi_{J^1Y_d} = 0$$

for all $\xi \in \mathfrak{g}$. Note that in the discrete case equivariance is the same as invariance, as the discrete Lagrangian is an approximation to the continuous action, rather than the continuous Lagrangian.

While we will not consider a general discrete momentum map for arbitrary actions, we define the vertical component to be the ***discrete Lagrangian momentum map*** $J_{L_d}^{E,X} : J^1Y_d \to \mathfrak{g}^*$ for an element E and base point X, which is

$$J_{L_d}^{E,X}(\xi) = \mathbf{i}_{\xi^V_{J^1Y_d}} \Theta_{L_d}^{E,X}.$$

We will now see that this is the appropriate definition for a discrete Noether's theorem.

Theorem 6.2 (Discrete Noether's theorem). *Consider a discrete Lagrangian system $L_d : J^1Y_d \to \mathbb{R}$ which is equivariant under the prolongation of the left action $\Phi : G \times Y_d \to Y_d$. Then the system satisfies the global conservation law*

$$\sum_{X \in \partial\mathcal{X}_d} \sum_{E \in \mathcal{E}_d(X)} J_{L_d}^{E,X}(\xi)(j^1\phi_d(E)) + D_H S_d(\phi_d) \cdot \pi_{\mathcal{X}}^{\partial} \cdot \xi_{\mathcal{C}_d(X)}(\phi_{d,\mathcal{X}}) = 0 \tag{6.40}$$

and the corresponding local conservation law

$$\sum_{X \in E} J_{L_d}^{E,X}(\xi)(j^1\phi_d(E)) + D_H L_d(\phi_d) \cdot \xi_{J^1Y_d}(j^1\phi_d(E)) = 0 \qquad \text{for all } E \in \mathcal{E}_d \tag{6.41}$$

for all solutions ϕ_d and all $\xi \in \mathfrak{g}$.

Proof. As we have already seen, the action of G on Y induces an action on Y_d and on J^1Y_d. This can then be extended to an action on the discrete configuration space $\mathcal{C}_d(Y)$. We use the equivariance of L_d to write

$$S_d(g \cdot \phi_d) = \sum_{E \in \mathcal{E}_d} L_d(g \cdot j^1\phi_d(E)) = \sum_{E \in \mathcal{E}_d} L_d(j^1\phi_d(E)) = S_d(\phi_d),$$

and so equivariance of the Lagrangian immediately implies that the action is also equivariant. Dif-

ferentiating this expression with respect to g gives

$$\mathbf{d}S_d(\phi_d) \cdot \xi_{\mathcal{C}_d(Y)}(\phi_d) = 0.$$

The group action thus maps solutions to solutions, and so $\xi_{\mathcal{C}_d(Y)}$ is tangent to the space of solutions $\mathcal{C}_{L_d}(Y)$. We can therefore use expression (6.38) to write the left-hand side of the previous equation as

$$\mathbf{d}S_d(\phi_d) \cdot \xi_{\mathcal{C}_d(Y)}(\phi_d) = \sum_{X \in \partial\mathcal{X}_d} \sum_{E \in \mathcal{E}_d(X)} \Theta_{L_d}^{E,X}(j^1\phi_d(E)) \cdot \xi_{J^1Y_d}^V(j^1\phi_d(E)) \\ + D_H S_d(\phi_d) \cdot \pi_{\mathcal{X}}^{\partial} \cdot \xi_{\mathcal{C}_d(X)}(\phi_{d,\mathcal{X}}),$$

and so equating our two expressions for the derivative of S_d in the group direction and using the definition of the discrete momentum map now gives the global statement of the discrete Noether's theorem. Taking the definition of infinitesimal equivariance of L_d and evaluating the left-hand side immediately gives the local statement. □ □

As in the continuous case, infinitesimal equivariance is sufficient for the discrete Noether's theorem to hold.

If we include the effects of boundary terms, as specified by (6.35), and we do not assume that the Lagrangian is equivariant (due to body forces, for example), then for arbitrary variations we have

$$\mathbf{d}S_d(\phi_d) \cdot V_d = \sum_{X \in \partial_1\mathcal{X}_d} \tau(X) \cdot V_{d,x_X} \\ + \sum_{X \in \partial\mathcal{X}_d \setminus \partial_1\mathcal{X}_d} \sum_{E \in \mathcal{E}_d(X)} \Theta_{L_d}^{E,X}(j^1\phi_d(E)) \cdot j^1V_d(E) + D_H S_d(\phi_d) \cdot V_{d,\mathcal{X}}^{\partial}.$$

If we now take the variation to be the infinitesimal symmetry action $V = \xi_{\mathcal{C}_d(Y)}$, then we obtain

$$\sum_{X \in \partial\mathcal{X}_d \setminus \partial_1\mathcal{X}_d} \sum_{E \in \mathcal{E}_d(X)} J_{L_d}^{E,X}(\xi)(j^1\phi_d(E)) = -\sum_{X \in \partial_1\mathcal{X}_d} \tau(X) \cdot \xi_Y(x_X) \\ - D_H S_d(\phi_d) \cdot \pi_{\mathcal{X}}^{\partial} \cdot \xi_{\mathcal{C}_d(X)}(\phi_{d,\mathcal{X}}) - \mathbf{d}S_d(\phi_d) \cdot \xi_{\mathcal{C}_d(Y)}(\phi_d). \quad (6.42)$$

This describes the extent to which the exact Noether conservation law is not satisfied due to boundary conditions and body forces, and is a discretization of (5.24).

Discrete symmetries and momentum maps. The three symmetry actions discussed in §5.2.6 are all linear, and so the linearity of the AVI discrete Lagrangian means that it inherits these symmetry groups as well. These then imply that linear momentum, angular momentum and energy

are preserved by the discrete system.

The calculations for linear and angular momentum for the AVI method are as presented in §6.1.4. Here the group acts vertically on the fibers of J^1Y_d, and the global form of Noether's theorem (6.40) gives whole-body conservation of linear and angular momentum.

For the time translation symmetry, the calculation reduces to the imposition of the horizontal Euler-Lagrange equation, as in §6.2.3. This then implies the local infinitesimal equivariance of the discrete Lagrangian and leads to whole-body conservation of energy.

In the case that there are traction boundary conditions or body forces, the exact Noether's theorem is not satisfied unless the tractions and body forces are zero in the infinitesimal symmetry directions. Instead we can use (6.42) to calculate the change in a whole-body conserved quantity due to the tractions.

6.4 Proof of AVI convergence

In this section we prove that the asynchronous variational integrators discussed in §6.1 converge as the time steps go to zero. We only consider convergence of the time stepping, and we take the spatial discretization to be fixed. This analysis applies equally well if AVI methods are applied directly to an ODE problem. In this section we denote time steps with the symbol h.

There are three key ideas needed for the proof:

1. The generalization of AVI methods to ***asynchronous splitting methods***, which provide a simple framework to consider convergence of the time stepping. This explicitly discards all information about the spatial discretization, focussing attention on the time stepping.

2. The formulation of a reasonable ***proxy system***, to which convergence can be proven. For standard integration techniques, after N steps of size h one proves that the integrator is close to the true flow map for time $T = Nh$. For asynchronous methods this is no longer correct, and in these notes we introduce equation (6.52) below as the proxy.

3. The condition under which convergence occurs. For standard time stepping methods we consider convergence as the timestep goes to zero. For asynchronous methods, however, it is possible for all time steps to go to zero but for the method to not converge to the true flow. We introduce the concept of ***maximum asynchronicity*** and we prove convergence as this tends to zero. For AVI methods this translates into a proof of convergence as the maximum time step goes to zero.

The technical assumptions made in this section are relatively standard. We work on unbounded Euclidean spaces $\mathbb{R}^n$ and assume that all vector fields are globally Lipschitz. Only autonomous systems are considered, as time dependancy can be included by adding a dimension with uniform

flow for time. We only consider non-negative time steps here, and leave the general case for future work.

6.4.1 Asychronous splitting methods (ASMs)

We will be concerned with systems on $\mathbb{R}^n$ of the form

$$\dot{x}(t) = f(x(t)) = \sum_{i=1}^{M} f_i(x(t)) \tag{6.43a}$$

$$x(0) = x_0, \tag{6.43b}$$

where the vector field f is the sum of M component vector fields f_i. We will denote by $\Phi^t : \mathbb{R}^n \to \mathbb{R}^n$ the flow map of f and by $\Phi_i^t : \mathbb{R}^n \to \mathbb{R}^n$ the flow map of f_i.

We assume that the component vector fields f_i are Lipschitz with constant L, so that

$$\|f_i(x) - f_i(y)\| \leq L\|x - y\| \quad \text{for } i = 1, \ldots, M \tag{6.44}$$

for all $x, y \in \mathbb{R}^n$. This implies that f is Lipschitz with constant ML. The flow maps Φ_i and Φ thus exist and are continuously differentiable for all t (Abraham et al. [1988]).

An asynchronous splitting method for the system (6.43) consists of M integrators $\Psi_i^h : \mathbb{R}^n \to \mathbb{R}^n$ for the component vector fields f_i together with a sequence of time steps $h_k \geq 0$ and indices i_k for $k = 1, \ldots, N$. The method is then defined by

$$y_k = \Psi_{i_k}^{h_k}(y_{k-1}) \quad \text{for } k = 1, \ldots, N \tag{6.45a}$$

$$y_0 = x_0 \tag{6.45b}$$

We define the cumulative time for component i to be

$$t_k^i = \sum_{j=1}^{k} \delta_{i_j}^i h_j \tag{6.46}$$

and the global minimum time as

$$t_k^{\min} = \min_{i=1,\ldots,M} t_k^i. \tag{6.47}$$

Observe that

$$t_{k+1}^{\min} - t_k^{\min} \leq h_{k+1} \quad \text{for } k = 1, \ldots, N-1. \tag{6.48}$$

The ***maximum asynchronicity*** is the smallest h such that

$$t_k^i - t_k^{\min} \leq h \quad \text{for } i = 1, \ldots, M,\ k = 1, \ldots, N \tag{6.49a}$$

$$h_k \leq h \quad \text{for } k = 1, \ldots, N. \tag{6.49b}$$

Note that this implies that

$$t_{k+1}^i - t_k^{\min} \leq h \quad \text{for } i = 1, \ldots, M,\ k = 1, \ldots, N-1. \tag{6.50}$$

We will assume that each integrator Ψ_i is consistent, so that there is a Lipschitz function $C_\Psi : \mathbb{R}^n \to \mathbb{R}$ with constant L_Ψ, and $h^* \in \mathbb{R}$ such that

$$\|\Psi_i^t(x) - \Phi_i^t(x)\| \leq t^2 C_\Psi(x) \quad \text{for all } t \leq h^* \tag{6.51}$$

for each $i = 1, \ldots, M$ and all $x \in \mathbb{R}^n$.

In common with standard synchronous splitting methods, asynchronous splitting methods are geometric in that conservation properties of the component Ψ_i are inherited by the overall integrator, as it is simply a composition. For example, if each Ψ_i is a symplectic map or if each Ψ_i preserves a quantity $C(x)$, then the complete method will also be symplectic or C preserving, respectively. Note that this does not imply that standard backward error analysis results (Reich [1999a], Hairer and Lubich [1997]) can be used to analyze the method, as these techniques rely on having repeated applications of a single map. While it may be possible to define a map $\bar{\Psi}$ so that $y_{k+1} = \bar{\Psi}(y_k)$ for some choices of sequences h_k and i_k, in general such a map does not exist.

The ***proxy system*** to which we will show convergence is defined by

$$z_k = \Phi^{t_k^{\min}}(x_0) + \sum_{i=1}^{M} (t_k^i - t_k^{\min}) f_m\Big(\Phi^{t_k^{\min}}(x_0)\Big) \tag{6.52a}$$

$$z_0 = x_0. \tag{6.52b}$$

Observe that as the maximum asynchronicity h tends to zero, the proxy system will converge to the true flow map Φ^t.

6.4.2 AVIs as ASMs

Asychronous variational integrators (AVIs) are a special case of asychronous splitting methods (ASMs). Here we use the notation from §6.1, extended to allow the possibility that the elemental time steps vary over time. Let Θ be the set of all elemental times, indexed sequentially so that $\Theta = \{t_a\}_{a=0}^{N_\Theta}$, and modified so that the initial time $t_1 = t_0$ only appears once. Thus $t_a \geq t_{a-1}$ for all

$a = 1, \ldots, N_\Theta$. Let $\{K_a\}_{a=1}^{N_\Theta}$ be the sequence of element indices and $\{\Delta t_a\}_{a=1}^{N_\Theta}$ be the sequence of element time steps, so that for each $a = 1, \ldots, N_\Theta$ there is a j with $2 \leq j \leq N_K$ such that $t_a = t_{K_a}^j$ and $\Delta t_a = t_{K_a}^j - t_{K_a}^{j-1}$.

The spatially discretized system for AVIs is given by

$$\begin{pmatrix} \dot{x} \\ \dot{p} \end{pmatrix} = f(x,p) = \begin{pmatrix} M^{-1}p \\ -\sum_{K \in \mathcal{T}} \frac{\partial V_K}{\partial x} \end{pmatrix}. \tag{6.53}$$

We consider the decomposition of f into $M = N_\mathcal{T} + 1$ vector fields f_i given by

$$f_i(x,p) = \begin{pmatrix} 0 \\ -\frac{\partial V_i}{\partial x} \end{pmatrix} \quad \text{for } i = 1, \ldots, N_\mathcal{T} \tag{6.54a}$$

$$f_{N_\mathcal{T}+1}(x,p) = \begin{pmatrix} M^{-1}p \\ 0 \end{pmatrix}. \tag{6.54b}$$

We now consider $N = 2N_\Theta$ time steps and indices specified by

$$h_{k-1} = t_a - t_{a-1} \qquad h_k = \Delta t_a \tag{6.55a}$$

$$i_{k-1} = N_\mathcal{T} + 1 \qquad i_k = K_a, \tag{6.55b}$$

where $k = 2a$ for $a = 1, \ldots, N_\Theta$. The method thus consists of alternate global "drift" and local "kick" steps.

The fact that the time sequence Θ is increasing means that the maximum asynchronicity h is equal to the maximum time step

$$h = \max_{K \in \mathcal{T}} \max_{j=2,\ldots,N_K} (t_K^j - t_K^{j-1}). \tag{6.56}$$

The maximum asynchronicity will thus tend to zero whenever the maximum elemental time step tends to zero.

6.4.3 Convergence proof

We first recall variants of Gronwall's inequality in discrete and continuous time.

Lemma 6.1. *Consider a sequence $\{e_k\}_{k=0}^N$ of real numbers satisfying*

$$e_k \leq A_k e_{k-1} + B_k \quad \text{for } k = 1, \ldots, N,$$

where $e_0 \geq 0$, $A_k \geq 1$ and $B_k \geq 0$ for all $k = 1, \ldots, N$. Then

$$e_k \leq \left(e_0 + \sum_{i=1}^{k} B_i \right) \prod_{i=1}^{k} A_i.$$

Lemma 6.2. *Consider a continuous real valued function $e(t)$ which satisfies*

$$e(t) \leq B + \int_0^t Ae(\tau)\ d\tau \quad \text{for } 0 \leq t \leq T$$

for some non-negative constants A, B, T. Then

$$e(t) \leq Be^{At} \quad \text{for } 0 \leq t \leq T.$$

Now we recall some standard bounds on flow maps of Lipschitz vector fields.

Lemma 6.3. *Given a vector field f on $\mathbb{R}^n$ which is globally Lipschitz with constant L and which has flow map Φ, then*

$$\|\Phi^t(x) - x\| \leq t\|f(x)\|e^{Lt} \tag{6.57}$$

$$\|\Phi^t(x) - \Phi^t(y)\| \leq \|x - y\|e^{Lt} \tag{6.58}$$

$$\left\|\Phi^t(x) - [x + tf(x)]\right\| \leq Lt^2\|f(x)\|e^{Lt} \tag{6.59}$$

for all $x, y \in \mathbb{R}^n$ and all $t \geq 0$.

Next we make precise the sense in which the proxy system is close to the true flow.

Lemma 6.4. *The proxy system satisfies*

$$\|\Phi^{t_k^{\min}}(x_0) - x_0\| \leq t_k^{\min} e^{MLt_k^{\min}} \sum_{i=1}^{M} \|f_i(x_0)\| \tag{6.60a}$$

$$\|z_k - \Phi^{t_k^{\min}}(x_0)\| \leq h(1 + MLt_k^{\min} e^{MLt_k^{\min}}) \sum_{i=1}^{M} \|f_i(x_0)\|. \tag{6.60b}$$

Proof. The bound (6.60a) follows directly from (6.57), which together with the definition (6.52a) then gives (6.60b). □

Now we are ready to prove convergence of asynchronous splitting methods. We derive a bound on the incremental error, using the decomposition illustrated in Figure 6.11.

Lemma 6.5. *The difference between the computed solution and the proxy system at time step $k+1$*

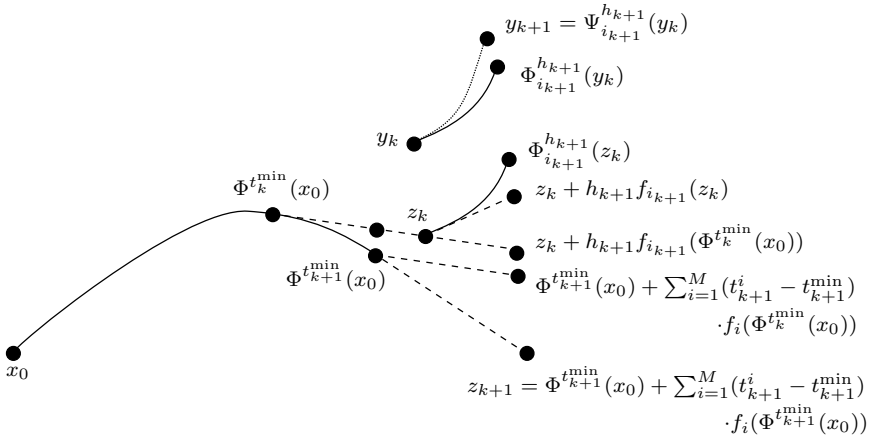


Figure 6.11: Decomposition of the error used in lemma 6.5.

satisfies

$$\|y_{k+1} - z_{k+1}\| \leq e^{(L+2L_\Psi)h_{k+1}}\|y_k - z_k\| + h_{k+1}h\left[C_\Psi(x_0) + C_B(L, L_\Psi, M, t_k^{\min})\sum_{i=1}^{M}\|f_i(x_0)\|\right], \quad (6.61)$$

where C_B is a smooth function which remains bounded as its arguments tend to zero.

Proof. Using the definitions of y_{k+1} and z_{k+1} we decompose the error at step $k+1$ as

$$\begin{aligned}
\|y_{k+1} - z_{k+1}\| \leq & \left\|\Psi_{i_{k+1}}^{h_{k+1}}(y_k) - \Phi_{i_{k+1}}^{h_{k+1}}(y_k)\right\| && (6.62a)\\
& + \left\|\Phi_{i_{k+1}}^{h_{k+1}}(y_k) - \Phi_{i_{k+1}}^{h_{k+1}}(z_k)\right\| && (6.62b)\\
& + \left\|\Phi_{i_{k+1}}^{h_{k+1}}(z_k) - \left[z_k + h_{k+1} f_{i_{k+1}}(z_k)\right]\right\| && (6.62c)\\
& + \Big\|\left[z_k + h_{k+1} f_{i_{k+1}}(z_k)\right] \\
& \qquad - \left[z_k + h_{k+1} f_{i_{k+1}}(\Phi^{t_k^{\min}}(x_0))\right]\Big\| && (6.62d)\\
& + \Big\|\left[z_k + h_{k+1} f_{i_{k+1}}(\Phi^{t_k^{\min}}(x_0))\right] \\
& \qquad - \left[\Phi^{t_{k+1}^{\min}}(x_0) + \sum_{i=1}^{M}(t_{k+1}^i - t_{k+1}^{\min}) f_i(\Phi^{t_k^{\min}}(x_0))\right]\Big\| && (6.62e)\\
& + \Big\|\left[\Phi^{t_{k+1}^{\min}}(x_0) + \sum_{i=1}^{M}(t_{k+1}^i - t_{k+1}^{\min}) f_i(\Phi^{t_k^{\min}}(x_0))\right] \\
& \qquad - \left[\Phi^{t_{k+1}^{\min}}(x_0) + \sum_{i=1}^{M}(t_{k+1}^i - t_{k+1}^{\min}) f_i(\Phi^{t_{k+1}^{\min}}(x_0))\right]\Big\|. && (6.62f)
\end{aligned}$$

Now (6.62e) is equal to

$$\left\| \left[\Phi^{t_k^{\min}}(x_0) + (t_{k+1}^{\min} - t_k^{\min}) f(\Phi^{t_k^{\min}}(x_0)) \right] - \Phi^{(t_{k+1}^{\min} - t_k^{\min})}(\Phi^{t_k^{\min}}(x_0)) \right\|$$

and (6.62f) is equal to

$$\left\| \sum_{i=1}^{M} (t_{k+1}^{i} - t_{k+1}^{\min}) \left[f_i(\Phi^{t_{k+1}^{\min}}(x_0)) - f_i(\Phi^{t_k^{\min}}(x_0)) \right] \right\|$$

and so using (6.51), (6.58), (6.59), (6.44), (6.49a) and (6.48) we obtain

$$\begin{aligned}
\|y_{k+1} - z_{k+1}\| \leq & h_{k+1}^2 C_\Psi(y_k) && \text{(6.63a)} \\
& + \|y_k - z_k\| e^{Lh_{k+1}} && \text{(6.63b)} \\
& + Lh_{k+1}^2 e^{Lh_{k+1}} \|f_{i_{k+1}}(z_k)\| && \text{(6.63c)} \\
& + h_{k+1} L \|z_k - \Phi^{t_k^{\min}}(x_0)\| && \text{(6.63d)} \\
& + MLhh_{k+1} e^{MLh_{k+1}} \|f(\Phi^{t_k^{\min}}(x_0))\| && \text{(6.63e)} \\
& + hh_{k+1} e^{Lh_{k+1}} \sum_{i=1}^{M} \|f_i(\Phi^{t_k^{\min}}(x_0))\|, && \text{(6.63f)}
\end{aligned}$$

where (6.63a)–(6.63f) are bounds for (6.62a)–(6.62f), respectively.

We now recall that for any Lipschitz function $a : \mathbb{R}^n \to \mathbb{R}^m$ with constant L_a, we have $\|a(y)\| \leq \|a(x)\| + L_a \|y - x\|$ for all $x, y \in \mathbb{R}^n$. Using this we can substitute the bounds in Lemma 6.4 into (6.63) to compute

$$\begin{aligned}
\|y_{k+1} - z_{k+1}\| \leq & (e^{Lh_{k+1}} + h_{k+1}^2 L_\Psi) \|y_k - z_k\| \\
& + h_{k+1} h \Bigg(C_\Psi(x_0) + \Bigg[L_\Psi h + L e^{Lh_{k+1}} + L^2 h e^{Lh_{k+1}} + e^{Lh_{k+1}} \\
& \quad + MLe^{MLh_{k+1}} + L + \Big(L_\Psi(hML + 1) + ML^2 \\
& \quad + L^2 e^{Lh_{k+1}}(hML + 1) + M^2 L^2 e^{MLh_{k+1}} \\
& \quad + MLe^{MLh_{k+1}} \Big) t_k^{\min} e^{MLt_k^{\min}} \Bigg] \sum_{i=1}^{M} \|f_i(x_0)\| \Bigg).
\end{aligned}$$

Finally, we note that $e^{Lh_{k+1}} + h_{k+1}^2 L_\Psi \leq e^{(L+2L_\Psi)h_{k+1}}$ and we use the restrictions $h \leq 1/L$ and $h \leq 1/L_\Psi$ to see that the function C_B depends only on L, L_Ψ, M and $t_k^{\min}$ and remains bounded as its arguments tend to zero. This gives the desired result. □

Lemma 6.6. *The difference between the computed solution and the proxy system at time step k*

satisfies

$$\|y_k - z_k\| \leq h(t_k^{\min} + h) M e^{(L+2L_\Psi)M(t_k^{\min}+h)} \Bigg[C_\Psi(x_0) + C_B(L, L_\Psi, M, t_k^{\min}) \sum_{i=1}^{M} \|f_i(x_0)\| \Bigg]. \quad (6.64)$$

Proof. This follows from applying lemma 6.1 to the estimate in lemma 6.5 and observing that from (6.46) and (6.49a) we have

$$\sum_{j=1}^{k} h_j = \sum_{i=1}^{M} t_k^i \leq \sum_{i=1}^{M} (t_k^{\min} + h) = M(t_k^{\min} + h).$$

□

Theorem 6.3. *Take a sequence of time step and index specifications $\{i_k^\alpha, h_k^\alpha\}_{k=1}^{N^\alpha}$ for $\alpha \in \mathbb{Z}^+$ with asychronicity bounds h^α. Assume that $h^\alpha \to 0$ and $t_N^{\min} \to T$ as $\alpha \to \infty$. Let $\{y_k\}_{k=1}^{N^\alpha}$ be the sequence generated by the asychronous splitting method with initial condition $y_0 = x_0$. Then*

$$\lim_{\alpha \to \infty} \|y_{N^\alpha} - x(T)\| = 0.$$

Proof. We decompose the global error to give

$$\|y_{N^\alpha} - x(T)\| \quad \leq \quad \|y_{N^\alpha} - z_{N^\alpha}\| + \|z_{N^\alpha} - \Phi^{t_{N^\alpha}^{\min}}(x_0)\| + \|\Phi^{t_{N^\alpha}^{\min}}(x_0) - \Phi^T(x_0)\|.$$

The three components each tend to zero as $\alpha \to \infty$, as can be seen from the bounds in lemmas 6.4 and 6.6, thus giving the result. □

Corollary 6.1 (Restatement of Theorem 6.1). *Consider a sequence of asychronous variational integrators for the same spatial discretization with maximum time step $h \to 0$ and maximum final time $t_{N_\Theta} \to T$. Then the final configuration converges to the exact solution $\Phi^T(x_0, p_0)$.*

Proof. We have that $t_{N_\Theta} - h \leq t_{N_\Theta}^{\min} \leq t_{N_\Theta}$ and so $t_{N_\Theta}^{\min} \to T$. The result then follows from theorem 6.3. □

Bibliography

R. Abraham, J. E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications*, volume 75 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition, 1988.

M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 1987.

H. Anderson. Rattle: A velocity version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 52:24–34, 1983.

F. Armero and I. Romero. On the formulation of high-frequency dissipative time-stepping algorithms for nonlinear dynamics. Part I: Low-order methods for two model problems and nonlinear elastodynamics. *Comput. Methods. Appl. Mech. Engrg.*, 190:2603–2649, 2001a.

F. Armero and I. Romero. On the formulation of high-frequency dissipative time-stepping algorithms for nonlinear dynamics. Part II: Second-order methods. *Comput. Methods. Appl. Mech. Engrg.*, 190:6783–6824, 2001b.

J. M. Arms, J. E. Marsden, and V. Moncrief. The structure of the space of solutions of Einstein's equations II : Several Killings fields and the Einstein-Yang-Mills equations. *Ann. of Phys.*, 144: 81–106, 1982.

U. M. Ascher and L. R. Petzold. *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM, 1998.

J. C. Baez and J. W. Gilliam. An algebraic approach to discrete mechanics. *Lett. Math. Phys.*, 31: 205–212, 1994.

E. Barth and B. Leimkuhler. Symplectic methods for conservative multibody systems. In *Integration Algorithms and Classical Mechanics (Toronto, ON 1993)*, pages 25–43. American Mathematical Society, 1996.

T. Belytschko. Partitioned and adaptive algorithms for explicit time integration. In W. Wunderlich, E. Stein, and K.-J. Bathe, editors, *Nonlinear Finite Element Analysis in Structural Mechanics*, pages 572–584. Springer-Verlag, 1981.

T. Belytschko and R. Mullen. Mesh partitions of explicit-implicit time integrators. In K.-J. Bathe, J. T. Oden, and W. Wunderlich, editors, *Formulations and Computational Algorithms in Finite Element Analysis*, pages 673–690. MIT Press, 1976.

G. Benettin and A. Giorgilli. On the Hamiltonian interpolation of near-to-the-identity symplectic mappings with application to symplectic integration algorithms. *J. Statist. Phys.*, 74:1117–1143, 1994.

J. S. Berg, R. L. Warnock, R. D. Ruth, and E. Forest. Construction of symplectic maps for nonlinear motion of particles in accelerators. *Physical Review E*, 49(1):722–739, 1994.

J. J. Biesiadecki and R. D. Skeel. Dangers of multiple time-step methods. *Journal of Computational Physics*, 109(2):318–328, 1993.

E. Binz, M. de Leon, D. M. de Diego, and D. Socolescu. Nonholonomic constraints in classical field theories. *Rep. Math. Phys.*, 49:151–166, 2002.

T. C. Bishop, R. D. Skeel, and K. Schulten. Difficulties with multiple time stepping and fast multipole algorithm in molecular dynamics. *Journal of Computational Chemistry*, 18(14):1785–1791, 1997.

A. M. Bloch, P. S. Krishnaprasad, J. E. Marsden, and T. S. Ratiu. The Euler-Poincaré equations and double bracket dissipation. *Comm. Math. Phys.*, 175:1–42, 1996.

A. I. Bobenko and Y. B. Suris. Discrete Lagrangian reduction, discrete Euler-Poincaré equations, and semidirect products. *Letters in Mathematical Physics*, 49(1):79–93, 1999a.

A. I. Bobenko and Y. B. Suris. Discrete time Lagrangian mechanics on Lie groups, with an application to the Lagrange top. *Communications in Mathematical Physics*, 204(1):147–188, 1999b.

M. Borri. Helicopter rotor dynamics by finite-element time approximation. *Computers & Mathematics with Applications-Part A*, 12(1):149–160, 1986.

C. L. Bottasso and O. A. Bauchau. Multibody modeling of engage and disengage operations of helicopter rotors. *J. Amer. Helic. Soc.*, 46, 2001.

C.L. Bottasso. A new look at finite elements in time: a variational interpretation of Runge-Kutta methods. *Applied Numerical Mathematics*, 25(4):355–368, 1997.

V. Brasey and E. Hairer. Symmetrized half-explicit methods for constrained mechanical systems. *Appl. Numer. Math.*, 13:23–31, 1993.

T. J. Bridges. Multi-symplectic structures and wave propagation. *Math. Proc. Camb. Phil. Soc.*, 121:147–90, 1997.

T. J. Bridges and G. Derks. Linear instability of solitary wave solutions of the Kawahara equation and its generalizations. *SIAM J. Math. Anal.*, 33:1356–1378, 2002.

T. J. Bridges and F. E. Laine-Pearson. Multisymplectic relative equilibria, multiphase wavetrains, and coupled NLS equations. *Stud. Appl. Math.*, 107:137–155, 2001.

T. J. Bridges and S. Reich. Multi-symplectic integrators: Numerical schemes for Hamiltonian PDEs that conserve symplecticity. *Physics Letters A*, 284(4-5):184–193, 2001a.

T. J. Bridges and S. Reich. Multi-symplectic spectral discretizations for the Zakharov-Kuznetsov and shallow water equations. *Physica D*, 152:491–504, 2001b.

J. A. Cadzow. Discrete calculus of variations. *Internat. J. Control.*, 11:393–407, 1970.

J. A. Cadzow. *Discrete-Time Systems: An Introduction with Interdisciplinary Applications*. Prentice-Hall, 1973.

B. Cano and R. Lewis. A comparison of symplectic and Hamilton's principle algorithms for autonomous and non-autonomous systems of ordinary differential equations. Technical report, Departamento de Matemática Aplicada y Computación, Universidad de Valladolid, 1998.

B. Cano and J. M. Sanz-Serna. Error growth in the numerical integration of periodic orbits, with application to Hamiltonian and reversible systems. *SIAM Journal on Numerical Analysis*, 34(4): 1391–1417, 1997.

P. J. Channell and C. Scovel. Symplectic integration of Hamiltonian systems. *Nonlinearity*, 3(2): 231–259, 1990.

J. B. Chen. New schemes for the nonlinear Schrödinger equation. *Appl. Math. Comput.*, 124:371–379, 2001.

J. B. Chen. Total variation in discrete multisymplectic field theory and multisymplectic-energy-momentum integrators. *Lett. Math. Phys.*, 61:63–73, 2002.

J. B. Chen. Multisymplectic geometry, local conservation laws and a multisymplectic integrator for the Zakharov-Kuznetsov equation. *Lett. Math. Phys.*, 63:115–124, 2003.

J. B. Chen and M. Z. Qin. A multisymplectic variational integrator for the nonlinear Schrödinger equation. *Numer. Meth. Part. Differ. Equ.*, 18:523–536, 2002.

F. Cirak and M. West. Decomposition-based Contact Response (DCR) for explicit dynamics. *International Journal for Numerical Methods in Engineering*, 2003. (submitted).

W. J. T. Daniel. Analysis and implementation of a new constant acceleration subcycling algorithm. *International Journal for Numerical Methods in Engineering*, 40:2841–2855, 1997a.

W. J. T. Daniel. The subcycled Newmark algorithm. *Computational Mechanics*, 20:272–281, 1997b.

R. De Vogelaére. Methods of integration which preserve the contact transformation property of the Hamiltonian equations. (University of Notre Dame preprint), 1956.

D. Estep and D. French. Global error control for the continuous Galerkin finite element method for ordinary differential equations. *Math. Mod. Numer. Anal.*, 28:815–852, 1994.

R. C. Fetecau, J. E. Marsden, M. Ortiz, and M. West. Nonsmooth Lagrangian mechanics. *SIAM Journal on Applied Dynamical Systems*, 2003a. (to appear).

R. C. Fetecau, J. E. Marsden, and M. West. Variational multisymplectic formulations of nonsmooth continuum mechanics. In E. Kaplan, J. E. Marsden, and K. R. Sreenivasan, editors, *Perspectives and Problems in Nonlinear Science*, pages 229–261. Springer Verlag, 2003b.

A. E. Fischer, J. E. Marsden, and V. Moncrief. The structure of the space of solutions of Einstein's equations, I: One Killing field. *Ann. Inst. H. Poincaré*, 33:147–194, 1980.

E. Forest and R. D. Ruth. 4th-order symplectic integration. *Physica D*, 43(1):105–117, 1990.

P.P. Friedmann. Numerical-methods for the treatment of periodic systems with applications to structural dynamics and helicopter rotor dynamics. *Computers & Structures*, 35(4):329–347, 1990.

Z. Ge and J. M. Marsden. Lie-Poisson integrators and Lie-Poisson Hamilton-Jacobi theory. *Phys. Lett. A*, 133:134–139, 1988.

S. Geng. Construction of high-order symplectic PRK methods. *J. Comput. Math.*, 13:40–50, 1995.

S. Geng. A simple way of constructing symplectic Runge-Kutta methods. *J. Comput. Math.*, 18: 61–68, 2000.

J. W. Gilliam. *Lagrangian and symplectic techniques in discrete mechanics.* PhD thesis, University of California, Riverside, Department of Mathematics, 1996. available from http://math.ucr.edu/home/baez.

R. Gillilan and K. Wilson. Shadowing, rare events and rubber bands. A variational Verlet algorithm for molecular dynamics. *J. Chem. Phys.*, 97(3):1757–1772, 1992.

H. Goldstein. *Classical Mechanics.* Addison-Wesley, second edition, 1980.

O. Gonzalez. *Design and analysis of conserving integrators for nonlinear Hamiltonian systems with symmetry.* PhD thesis, Stanford University, Department of Mechanical Engineering, 1996a.

O. Gonzalez. Time integration and discrete Hamiltonian systems. *Journal of Nonlinear Science*, 6 (5):449–467, 1996b.

O. Gonzalez. Mechanical systems subject to holonomic constraints: Differential-algebraic formulations and conservative integration. *Physica D*, 132(1-2):165–174, 1999.

O. Gonzalez, D. J. Higham, and A. M. Stuart. Qualitative properties of modified equations. *IMA Journal of Numerical Analysis*, 19(2):169–190, 1999.

O. Gonzalez and J. C. Simo. On the stability of symplectic and energy-momentum algorithms for non-linear Hamiltonian systems with symmetry. *Computer Methods in Applied Mechanics and Engineering*, 134(3-4):197–222, 1996.

M. Gotay, J. Isenberg, and J. E. Marsden. Momentum maps and classical relativistic fields, part I: Covariant field theory. (unpublished), 1997.

H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Sim.*, 6:121–142, 1991.

H. Y. Guo, X. M. Ji, Y. Q. Li, and K. Wu. A note on symplectic, multisymplectic scheme in finite element method. *Commun. Theor. Phys.*, 36:259–262, 2001a.

H. Y. Guo, Y. Q. Li, and K. Wu. On symplectic and multisymplectic structures and their discrete versions in Lagrangian formalism. *Commun. Theor. Phys.*, 35:703–710, 2001b.

H. Y. Guo, Y. Q. Li, K. Wu, and S. K. Wang. Difference discrete variational principles, Euler-Lagrange cohomology and symplectic, multisymplectic structures I: Difference discrete variational principle. *Commun. Theor. Phys.*, 37:1–10, 2002a.

H. Y. Guo, Y. Q. Li, K. Wu, and S. K. Wang. Difference discrete variational principles, Euler-Lagrange cohomology and symplectic, multisymplectic structures II: Euler-Lagrange cohomology. *Commun. Theor. Phys.*, 37:129–138, 2002b.

H. Y. Guo, Y. Q. Li, K. Wu, and S. K. Wang. Difference discrete variational principles, Euler-Lagrange cohomology and symplectic, multisymplectic structures III: Application to symplectic and multisymplectic algorithms. *Commun. Theor. Phys.*, 37:257–264, 2002c.

M. E. Gurtin. *Configurational forces as basic concepts of continuum physics*. Springer, 2000.

E. Hairer. Backward analysis of numerical integrators and symplectic methods. *Annals of Numerical Mathematics*, 1:107–132, 1994.

E. Hairer. Symmetric projection methods for differential equations on manifolds. *BIT*, 40(4):726–734, 2000.

E. Hairer and C. Lubich. The life-span of backward error analysis for numerical integrators. *Numerische Mathematik*, 76(4):441–462, 1997.

E. Hairer and C. Lubich. Invariant tori of dissipatively perturbed Hamiltonian systems under symplectic discretization. *Applied Numerical Mathematics*, 29(1):57–71, 1999.

E. Hairer and C. Lubich. Long-time energy conservation of numerical methods for oscillatory differential equations. *SIAM Journal on Numerical Analysis*, 38(2):414–441, 2000.

E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, 2002.

E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I : Nonstiff problems*, volume 8 of *Springer Series in Computational Mathematics*. Springer-Verlag, second edition, 1993.

E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II : Stiff and differential-algebraic problems*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, second edition, 1996.

W. R. Hamilton. On a general method in dynamics. *Philos. Trans. Royal Soc. London*, 1834. Part II, 247–308; Part I for 1835, 95–144.

D. J. Hardy, D. I. Okunbor, and R. D. Skeel. Symplectic variable step size integration for N-body problems. *Applied Numerical Mathematics*, 29(1):19–30, 1999.

J. L. Hong and M. Z. Qin. Multisymplecticity of the centred box discretization for Hamiltonian PDEs with $m \geq 2$ space dimensions. *Appl. Math. Lett.*, 15:1005–1011, 2002.

T. J. R Hughes. *The Finite Element Method : Linear Static and Dynamic Finite Element Analysis*. Prentice-Hall, 1987.

T. J. R. Hughes and W. K. Liu. Implicit-explicit finite elements in transient analysis: Stability theory. *Journal of Applied Mechanics*, 78:371–374, 1978.

T. J. R. Hughes, K. S. Pister, and R. L. Taylor. Implicit-explicit finite elements in nonlinear transient analysis. *Computer Methods In Applied Mechanics And Engineering*, 17/18:159–182, 1979.

B. Hulme. Discrete Galerkin and related one-step methods for ordinary differential equations. *Math. Comp.*, 26:881–891, 1972a.

B. Hulme. One-step piecewise polynomial Galerkin methods for initial value problems. *Math. Comp.*, 26:415–426, 1972b.

C. L. Hwang and L. T. Fan. A discrete version of Pontryagin's maximum principle. *Operations Res.*, 15:139–146, 1967.

P. E. Hydon. Conservation laws of partial difference equations with two independent variables. *J. Phys. A-Math. Gen.*, 34:10347–10355, 2001.

A. L. Islas, D. A. Karpeev, and C. M. Schober. Geometric integrators for the nonlinear Schrödinger equation. *J. Comput. Phys.*, 173:116–148, 2001.

A. L. Islas and C. M. Schober. Multisymplectic spectral methods for the Gross-Pitaevskii equation. *Lect. Note. Comput. Sci.*, 2331:486–495, 2002.

T. Itoh and K. Abe. Hamiltonian-conserving discrete canonical equations based on variational difference equations. *Journal of Computational Physics*, 77:85–102, 1988.

J. A. Izaguirre, S. Reich, and R. D. Skeel. Longer time steps for molecular dynamics. *Journal of Chemical Physics*, 110(20):9853–9864, 1999.

C. G. K. Jacobi. *Vorlesungen über Dynamik*. Verlag G. Reimer, 1866.

S. M. Jalnapurkar, M. Leok, J. E. Marsden, and M. West. Discrete Routh reduction. *Foundations of Computational Mathematics*, 2003. (to appear).

G. Jaroszkiewicz and K. Norton. Principles of discrete time mechanics, I: Particle systems. *J. Phys. A*, 30:3115–3144, 1997a.

G. Jaroszkiewicz and K. Norton. Principles of discrete time mechanics, II: Classical field theory. *J. Phys. A*, 30:3145–3163, 1997b.

L. O. Jay. Symplectic partitioned Runge-Kutta methods for constrained Hamiltonian systems. *SIAM Journal on Numerical Analysis*, 33(1):368–387, 1996.

L. O. Jay. Structure preservation for constrained dynamics with super partitioned additive Runge-Kutta methods. *SIAM Journal on Scientific Computing*, 20(2):416–446, 1999.

C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Cambridge University Press, 1987.

B. W. Jordan and E. Polak. Theory of a class of discrete optimal control systems. *J. Electronics Control*, 17:697–711, 1964.

C. Kane, J. E. Marsden, and M. Ortiz. Symplectic-energy-momentum preserving variational integrators. *Journal of Mathematical Physics*, 40(7):3353–3371, 1999a.

C. Kane, J. E. Marsden, M. Ortiz, and M. West. Variational integrators and the Newmark algorithm for conservative and dissipative mechanical systems. *International Journal for Numerical Methods in Engineering*, 49(10):1295–1325, 2000.

C. Kane, E. A. Repetto, M. Ortiz, and J. E. Marsden. Finite element analysis of nonsmooth contact. *Computer Methods in Applied Mechanics and Engineering*, 180(1-2):1–26, 1999b.

J. Kijowski and W. Tulczyjew. *A Symplectic Framework for Field Theories*, volume 107 of *Lecture Notes in Physics*. Springer, 1979.

D. Knuth. *The Art of Computer Programming*. Addisson-Wesley, 1998.

S. Kouranbaeva and S. Shkoller. A variational approach to second-order multisymplectic field theory. *J. Geom. Phys.*, 35(4):333–366, 2000.

R. A. Labudde and D. Greenspan. Discrete mechanics—A general treatment. *Journal of Computational Physics*, 15:134–167, 1974.

R. A. Labudde and D. Greenspan. Energy and momentum conserving methods of arbitrary order for the numerical integration of equations of motion—I. Motion of a single particle. *Numer. Math.*, 25:323–346, 1976a.

R. A. Labudde and D. Greenspan. Energy and momentum conserving methods of arbitrary order for the numerical integration of equations of motion—II. Motion of a system of particles. *Numer. Math.*, 26:1–16, 1976b.

S. Lall and M. West. Discrete variational mechanics and duality. (in preparation), 2003.

F. M. Lasagni. Canonical Runge-Kutta methods. *ZAMP*, 39:952–953, 1988.

T. D. Lee. Can time be a discrete dynamical variable? *Phys. Lett. B*, 122:217–220, 1983.

T. D. Lee. Difference equations and conservation laws. *J. Stat. Phys.*, 46:843–860, 1987.

B. Leimkuhler and G. W. Patrick. A symplectic integrator for Riemannian manifolds. *Journal of Nonlinear Science*, 6(4):367–384, 1996.

B. Leimkuhler and S. Reich. Symplectic integration of constrained Hamiltonian systems. *Mathematics of Computation*, 63(208):589–605, 1994.

B. J. Leimkuhler and R. D. Skeel. Symplectic numerical integrators in constrained Hamiltonian-systems. *Journal of Computational Physics*, 112(1):117–125, 1994.

A. Lew, J. E. Marsden, M. Ortiz, and M. West. Asynchronous variational integrators. *Archive for Rational Mechanics and Analysis*, 167(2):85–146, 2003a.

A. Lew, J. E. Marsden, M. Ortiz, and M. West. Variational time integrators. *International Journal for Numerical Methods in Engineering*, 2003b. (to appear).

T. T. Liu and M. Z. Qin. Multisymplectic geometry and multisymplectic Preissman scheme for the KP equation. *J. Math. Phys.*, 43:4060–4077, 2002.

J. D. Logan. First integrals in the discrete calculus of variations. *Aequationes Mathematicae*, 9: 210–220, 1973.

P. S. Krishnaprasad M. A. Austin and L. S. Wang. Almost Poisson integration of rigid body systems. *J. Comput. Phys.*, 107:105–117, 1993.

R. MacKay. Some aspects of the dynamics of Hamiltonian systems. In D. S. Broomhead and A. Iserles, editors, *The dynamics of numerics and the numerics of dynamics*, pages 137–193. Clarendon Press, Oxford, 1992.

S. Maeda. Canonical structure and symmetries for discrete systems. *Math. Japonica*, 25:405–420, 1980.

S. Maeda. Extension of discrete Noether theorem. *Math. Japonica*, 26:85–90, 1981a.

S. Maeda. Lagrangian formulation of discrete systems and concept of difference space. *Math. Japonica*, 27:345–356, 1981b.

J. E. Marsden. Park city lectures on mechanics, dynamics and symmetry. In Y. Eliashberg and L. Traynor, editors, *Symplectic Geometry and Topology*, volume 7 of *IAS/Park City Math. Ser.*, pages 335–430. American Mathematical Society, 1999.

J. E. Marsden and T. J. R. Hughes. *Mathematical Foundations of Elasticity*. Dover Publications, 1994.

J. E. Marsden, G. W. Patrick, and S. Shkoller. Multisymplectic geometry, variational integrators, and nonlinear PDEs. *Communications in Mathematical Physics*, 199(2):351–395, 1998.

J. E. Marsden, S. Pekarsky, and S. Shkoller. Discrete Euler-Poincaré and Lie-Poisson equations. *Nonlinearity*, 12(6):1647–1662, 1999a.

J. E. Marsden, S. Pekarsky, and S. Shkoller. Stability of relative equilibria of point vortices on a sphere and symplectic integrators. *Nuovo Cimento Della Societa Italiana Di Fisica C—Geophysics and Space Physics*, 22(6):793–802, 1999b.

J. E. Marsden, S. Pekarsky, S. Shkoller, and M. West. Variational methods, multisymplectic geometry and continuum mechanics. *Journal of Geometry and Physics*, 38(3-4):253–284, 2001.

J. E. Marsden and T. Ratiu. *Introduction to Mechanics and Symmetry*, volume 17 of *Texts in Applied Mathematics*. Springer-Verlag, second edition, 1999.

J. E. Marsden and S. Shkoller. Multisymplectic geometry, covariant Hamiltonians, and water waves. *Mathematical Proceedings of the Cambridge Philosophical Society*, 125(3):553–575, 1999.

J. E. Marsden and M. West. Discrete mechanics and variational integrators. In *Acta Numerica*, volume 10. Cambridge University Press, 2001.

R. I. McLachlan. On the numerical integration of ordinary differential equations by symmetric composition methods. *SIAM J. Sci. Comp.*, 16:151–168, 1993.

R. I. McLachlan, G. R. W. Quispel, and N. Robidoux. Unified approach to Hamiltonian systems, Poisson systems, gradient systems, and systems with Lyapunov functions or first integrals. *Physical Review Letters*, 81(12):2399–2403, 1998.

R. I. McLachlan, G. R. W. Quispel, and N. Robidoux. Geometric integration using discrete gradients. *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, 357(1754):1021–1045, 1999.

R. I. McLachlan and C. Scovel. Equivariant constrained symplectic integration. *Journal of Nonlinear Science*, 5(3):233–256, 1995.

J. Moser and A. P. Veselov. Discrete versions of some classical integrable systems and factorization of matrix polynomials. *Communications in Mathematical Physics*, 139(2):217–243, 1991.

S. Müller and M. Ortiz. On the Γ-convergence of discrete dynamics and variational integrators. (preprint), 2003.

A. Murua and J. M. Sanz-Serna. Order conditions for numerical integrators obtained by composing simpler integrators. *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, 357(1754):1079–1100, 1999.

U. Mutze. Predicting classical motion directly from the action principle. (preprint), 1998.

M. O. Neal and T. Belytschko. Explicit-explicit subcycling with non-integer time step ratios for structural dynamic systems. *Computers & Structures*, 6:871–880, 1989.

N. Newmark. A method of computation for structural dynamics. *ASCE Journal of the Engineering Mechanics Division*, 85(EM 3):67–94, 1959.

E. Noether. Invariante variationsprobleme. *Kgl. Ges. Wiss. Nachr. Göttingen. Math. Physik*, 2: 235–257, 1918.

K. Norton and G. Jaroszkiewicz. Principles of discrete time mechanics, III: Quantum field theory. *J. Phys. A*, 31:977–1000, 1998.

M. Oliver, M. West, and C. Wulff. Approximate momentum conservation for spatial semidiscretizations of semilinear wave equations. *Numerische Mathematik*, 2003. (to appear).

P. J. Oliver and J. Sivaloganathan. The structure of null Lagrangians. *Nonlinearity*, 1:389–398, 1988.

M. Ortiz and L. Stainier. The variational formulation of viscoplastic constitutive updates. *Computer Methods in Applied Mechanics and Engineering*, 171(3-4):419–444, 1999.

S. Pekarsky and M. West. Discrete diffeomorphism groupoids and circulation conserving fluid integrators. (in preparation), 2003.

M. Qin and W. J. Zhu. Construction of higher order symplectic schemes by composition. *Computing*, 27:309–321, 1992.

R. Radovitzky and M. Ortiz. Error estimation and adaptive meshing in strongly nonlinear dynamic problems. *Computer Methods in Applied Mechanics and Engineering*, 172:203–240, 1999.

S. Reich. Symplectic integration of constrained Hamiltonian systems by composition methods. *SIAM Journal on Numerical Analysis*, 33(2):475–491, 1996.

S. Reich. On higher-order semi-explicit symplectic partitioned Runge-Kutta methods for constrained Hamiltonian systems. *Numerische Mathematik*, 76(2):231–247, 1997.

S. Reich. Backward error analysis for numerical integrators. *SIAM Journal on Numerical Analysis*, 36(5):1549–1570, 1999a.

S. Reich. Multiple time scales in classical and quantum-classical molecular dynamics. *Journal of Computational Physics*, 151(1):49–73, 1999b.

S. Reich. Finite volume methods for multi-symplectic PDEs. *BIT*, 40(3):559–582, 2000a.

S. Reich. Multi-symplectic Runge-Kutta collocation methods for Hamiltonian wave equations. *Journal of Computational Physics*, 157(2):473–499, 2000b.

C. W. Rowley and J. E. Marsden. Variational integrators for point vortices. *Proc. CDC*, 40, 2002.

R. D. Ruth. A canonical integration technique. *IEEE Transactions on Nuclear Science*, 30(4): 2669–2671, 1983.

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. *Journal of Computational Physics*, 23:327–341, 1977.

J. M. Sanz-Serna. Runge-Kutta schemes for Hamiltonian systems. *BIT*, 28(4):877–883, 1988.

J. M. Sanz-Serna. The numerical integration of Hamiltonian systems. In J. R. Cash and I. Gladwell, editors, *Computional ordinary differential equations*, pages 437–449. Clarendon Press, Oxford, 1992a.

J. M. Sanz-Serna. Symplectic Runge-Kutta and related methods—Recent results. *Physica D*, 60 (1-4):293–302, 1992b.

J. M. Sanz-Serna and M. P. Calvo. *Numerical Hamiltonian Problems*. Chapman and Hall, 1994.

T. Schlick, R. D. Skeel, A. T. Brunger, L. V. Kale, J. A. Board, J. Hermans, and K. Schulten. Algorithmic challenges in computational molecular biophysics. *Journal of Computational Physics*, 151(1):9–48, 1999.

W. M. Seiler. Numerical analysis of constrained Hamiltonian systems and the formal theory of differential equations. *Mathematics and Computers in Simulation*, 45(5-6):561–576, 1998a.

W. M. Seiler. Position versus momentum projections for constrained Hamiltonian systems. *Numerical Algorithms*, 19(1-4):223–234, 1998b.

W. M. Seiler. Numerical integration of constrained Hamiltonian systems using Dirac brackets. *Mathematics of Computation*, 68(226):661–681, 1999.

G. Sheng, T. C. Fung, and S. C. Fan. Parametrized formulations of Hamilton's law for numerical solutions of dynamic problems: Part II. Time finite element approximation. *Computational Mechanics*, 21(6):449–460, 1998.

Y. Shibberu. Time-discretization of Hamiltonian systems. *Computers Math. Applic.*, 28(10–12): 123–145, 1994.

M. Shimada and H. Yoshida. Long-term conservation of adiabatic invariants by using symplectic integrators. *Publ. Astronomical Soc. Japan*, 48:147–155, 1996.

J. C. Simo and N. Tarnow. The discrete energy-momentum method—Conserving algorithms for nonlinear elastodynamics. *Zeitschrift für Angewandte Mathematik und Physik*, 43(5):757–792, 1992.

J. C. Simo, N. Tarnow, and K. K. Wong. Exact energy-momentum conserving algorithms and symplectic schemes for nonlinear dynamics. *Computer Methods in Applied Mechanics and Engineering*, 100(1):63–116, 1992.

R. D. Skeel and K. Srinivas. Nonlinear stability analysis of area-preserving integrators. *SIAM Journal on Numerical Analysis*, 38(1):129–148, 2000.

R. D. Skeel, G. H. Zhang, and T. Schlick. A family of symplectic integrators: Stability, accuracy, and molecular dynamics applications. *SIAM Journal on Scientific Computing*, 18(1):203–222, 1997.

P. Smolinski and Y.-S. Wu. An implicit multi-time step integration method for structural dynamics problems. *Computational Mechanics*, 1998.

M. Sofroniou and W. Oevel. Symplectic Runge-Kutta-schemes I: Order conditions. *SIAM Journal of Numerical Analysis*, 34(5):2063–2086, 1997a.

M. Sofroniou and W. Oevel. Symplectic Runge-Kutta-schemes II: Classification of symmetric methods. preprint, 1997b.

Y. J. Sun and M. Z. Qin. Construction of multisymplectic schemes of any finite order for modified wave equations. *J. Math. Phys.*, 41:7854–7868, 2000.

Y. Suris. Hamiltonian methods of Runge-Kutta type and their variational interpretation. *Mathematical Simulation*, 2(4):78–87, 1990.

Y. B. Suris. The canonicity of mappings generated by Runge-Kutta type methods when integrating the system $\ddot{x} = -\partial u/\partial x$. *USSR Computational Mathematics and Mathematical Physics*, 29(1): 138–144, 1989.

W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer-simulation method for the calculation of equilibrium-constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.*, 76:637–649, 1982.

V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*. Springer-Verlag, New-York, 1997.

C. Truesdell and W. Noll. The non-linear field theories of mechanics. In S. Flugge, editor, *Handbuch der Physik*, volume III/3. Springer-Verlag, 1965.

M. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, 97:1990–2001, 1992.

L. Verlet. Computer experiments on classical fluids. *Phys. Rev.*, 159:98–103, 1967.

A. P. Veselov. Integrable discrete-time systems and difference operators. *Functional Analysis and its Applications*, 22(2):83–93, 1988.

A. P. Veselov. Integrable Lagrangian correspondences and the factorization of matrix polynomials. *Functional Analysis and its Applications*, 25(2):112–122, 1991.

Y. S. Wang and M. Z. Qin. Multisymplectic geometry and multisymplectic scheme for the nonlinear Klein-Gordon equation. *J. Phys. Soc. Jpn.*, 70:653–661, 2001.

Y. S. Wang and M. Z. Qin. Multisymplectic schemes for the nonlinear Klein-Gordon equation. *Math. Comput. Model.*, 36:963–977, 2002.

R. L. Warnock and R. D. Ruth. Stability of nonlinear Hamiltonian motion for a finite but very long-time. *Physical Review Letters*, 66(8):990–993, 1991.

R. L. Warnock and R. D. Ruth. Long-term bounds on nonlinear Hamiltonian motion. *Physica D*, 56(2-3):188–215, 1992.

J. M. Wendlandt and J. E. Marsden. Mechanical integrators derived from a discrete variational principle. *Physica D*, 106(3-4):223–246, 1997a.

J. M. Wendlandt and J. E. Marsden. Mechanical systems with symmetry, variational principles and integration algorithms. In M. Alber, B. Hu, and J. Rosenthal, editors, *Current and Future Directions in Applied Mathematics*, pages 219–261. Birkhäuser, 1997b.

J. Wisdom, S. J. Peale, and F. Mignard. The chaotic rotation of Hyperion. *Icarus*, 58:137–152, 1984.

H. Yoshida. Construction of higher-order symplectic integrators. *Physics Letters A*, 150(5-7):262–268, 1990.