Article

R-adaptive multisymplectic and variational integrators

Tomasz M. Tyranowski [1,2,‡*] and Mathieu Desbrun [2,‡]

1 Max-Planck-Institut für Plasmaphysik, Boltzmannstraße 2, 85748 Garching, Germany
2 California Institute of Technology, Computing + Mathematical Sciences, Pasadena, CA 91125, USA
* Correspondence: tomasz.tyranowski@ipp.mpg.de
‡ These authors contributed equally to this work.

Version July 18, 2019 submitted to Mathematics

Abstract: Moving mesh methods (also called *r*-adaptive methods) are space-adaptive strategies used for the numerical simulation of time-dependent partial differential equations. These methods keep the total number of mesh points fixed during the simulation, but redistribute them over time to follow the areas where a higher mesh point density is required. There are a very limited number of moving mesh methods designed for solving field-theoretic partial differential equations, and the numerical analysis of the resulting schemes is challenging. In this paper we present two ways to construct *r*-adaptive variational and multisymplectic integrators for (1+1)-dimensional Lagrangian field theories. The first method uses a variational discretization of the physical equations and the mesh equations are then coupled in a way typical of the existing *r*-adaptive schemes. The second method treats the mesh points as pseudo-particles and incorporates their dynamics directly into the variational principle. A user-specified adaptation strategy is then enforced through Lagrange multipliers as a constraint on the dynamics of both the physical field and the mesh points. We discuss the advantages and limitations of our methods. Numerical results for the Sine-Gordon equation are also presented.

Keywords: geometric numerical integration; variational integrators; multisymplectic integrators; field theory; moving mesh methods; moving mesh partial differential equations; solitons; Sine-Gordon equation

---

1. Introduction

The purpose of this work is to design, analyze and implement variational and multisymplectic integrators for Lagrangian partial differential equations with space-adaptive meshes. In this paper we combine geometric numerical integration and *r*-adaptive methods for the numerical solution of PDEs. We show that these two fields are compatible, mostly due to the fact that in *r*-adaptation the number of mesh points remains constant and we can treat them as additional pseudo-particles whose dynamics is coupled to the dynamics of the physical field of interest.

Geometric (or structure-preserving) integrators are numerical methods that preserve geometric properties of the flow of a differential equation (see [1]). This encompasses symplectic integrators for Hamiltonian systems, variational integrators for Lagrangian systems, and numerical methods on manifolds, including Lie group methods and integrators for constrained mechanical systems. Geometric integrators proved to be extremely useful for numerical computations in astronomy, molecular dynamics, mechanics and theoretical physics. The main motivation for developing structure-preserving algorithms lies in the fact that they show excellent numerical behavior, especially for long-time integration of equations possessing geometric properties.

An important class of structure-preserving integrators are *variational integrators* for Lagrangian systems ([1], [2]). This type of integrator is based on discrete variational principles. The variational approach provides a unified framework for the analysis of many symplectic algorithms and is characterized by a natural treatment of the discrete Noether theorem, as well as forced, dissipative and constrained systems. Variational integrators were first introduced in the context of finite-dimensional mechanical systems, but later Marsden, Patrick and Shkoller [3] generalized this idea to field theories. Variational integrators have since then been successfully applied in many computations, for example in elasticity ([4]), electrodynamics ([5]) or fluid dynamics ([6]). Existing variational integrators so far have been developed on static, mostly uniform spatial meshes. The main goal of this paper is to design and analyze variational integrators that allow for the use of space-adaptive meshes.

Adaptive meshes used for the numerical solution of partial differential equations fall into three main categories: h-adaptive, p-adaptive and r-adaptive. R-adaptive methods, which are also known as *moving mesh methods* ([7], [8]), keep the total number of mesh points fixed during the simulation, but relocate them over time. These methods are designed to minimize the error of the computations by optimally distributing the mesh points, contrasting with h-adaptive methods for which the accuracy of the computations is obtained via insertion and deletion of mesh points. Moving mesh methods are a large and interesting research field of applied mathematics, and their role in modern computational modeling is growing. Despite the increasing interest in these methods in recent years, they are still in a relatively early stage of their development compared to the more matured h-adaptive methods.

Overview

There are three logical steps to r-adaptation:

- Discretization of the physical PDE
- Mesh adaptation strategy
- Coupling the mesh equations to the physical equations

The key ideas of this paper regard the first and the last step. Following the general spirit of variational integrators, we discretize the underlying action functional rather than the PDE itself, and then derive the discrete equations of motion. We base our adaptation strategies on the equidistribution principle and the resulting moving mesh partial differential equations (MMPDEs). We interpret MMPDEs as constraints, which allows us to consider *novel* ways of coupling them to the physical equations. Note that we will restrict our explanations to one time and one space dimension for the sake of simplicity.

Let us consider a (1+1)-dimensional scalar field theory with the action functional

$$S[\phi] = \int_0^{T_{max}} \int_0^{X_{max}} \mathcal{L}(\phi, \phi_X, \phi_t)\, dX\, dt, \tag{1}$$

where $\phi : [0, X_{max}] \times [0, T_{max}] \longrightarrow \mathbb{R}$ is the field and $\mathcal{L} : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ its Lagrangian density. For simplicity, we assume the following fixed boundary conditions

$$\begin{aligned} \phi(0, t) &= \phi_L, \\ \phi(X_{max}, t) &= \phi_R. \end{aligned} \tag{2}$$

In order to further consider moving meshes let us perform a change of variables $X = X(x, t)$ such that for all t the map $X(., t) : [0, X_{max}] \longrightarrow [0, X_{max}]$ is a 'diffeomorphism'—more precisely, we only require that $X(., t)$ is a homeomorphism such that both $X(., t)$ and $X(., t)^{-1}$ are piecewise C^1. In the context of mesh adaptation the map $X(x, t)$ represents the spatial position at time t of the mesh point labeled by x. Define $\varphi(x, t) = \phi(X(x, t), t)$. Then the partial derivatives of ϕ are $\phi_X(X(x, t), t) = \varphi_x / X_x$ and $\phi_t(X(x, t), t) = \varphi_t - \varphi_x X_t / X_x$. Plugging these equations in (1) we get

$$S[\phi] = \int_0^{T_{max}} \int_0^{X_{max}} \mathcal{L}\Big(\varphi, \frac{\varphi_x}{X_x}, \varphi_t - \frac{\varphi_x X_t}{X_x}\Big) X_x \, dx \, dt =: \tilde{S}[\varphi], \tilde{S}[\varphi, X] \tag{3}$$

where the last equality defines two modified, or 'reparametrized', action functionals. For the first one, $\tilde{S}$ is considered as a functional of φ only, whereas in the second one we also treat it as a functional of X. This leads to two different approaches to mesh adaptation, which we dub the *control-theoretic* strategy and the *Lagrange multiplier* strategy, respectively. The 'reparametrized' field theories defined by $\tilde{S}[\varphi]$ and $\tilde{S}[\varphi, X]$ are both intrinsically covariant; however, it is convenient for computational purposes to work with a space-time split and formulate the field dynamics as an initial value problem.

Outline

This paper is organized as follows. In Section 2 and Section 3 we take the view of infinite dimensional manifolds of fields as configuration spaces, and develop the control-theoretic and Lagrange multiplier strategies in that setting. It allows us to discretize our system in space first and consider time discretization later on. It is clear from our exposition that the resulting integrators are variational. In Section 4 we show how similar integrators can be constructed using the covariant formalism of multisymplectic field theory. We also show how the integrators from the previous sections can be interpreted as multisymplectic. In Section 5 we apply our integrators to the Sine-Gordon equation and we present our numerical results. We summarize our work in Section 6 and discuss several directions in which it can be extended.

2. Control-theoretic approach to r-adaptation

At first glance, it appears that the simplest and most straightforward way to construct an r-adaptive variational integrator would be to discretize the physical system in a similar manner to the general approach to variational integration, i.e. discretize the underlying variational principle, and then derive the mesh equations and couple them to the physical equations in a way typical of the existing r-adaptive algorithms. We explore this idea in this section and show that it indeed leads to space adaptive integrators that are variational in nature. However, we also show that those integrators do not exhibit the behavior expected of geometric integrators, such as good energy conservation. We will refer to this strategy as *control-theoretic*, since in this description the field φ represents the physical state of the system, while X can be interpreted as a control variable and the mesh equations as feedback (see, e.g., [9]).

2.1. Reparametrized Lagrangian

For the moment let us assume that $X(x,t)$ is a known function. We denote by $\xi(X,t)$ the function such that $\xi(.,t) = X(.,t)^{-1}$, that is $\xi(X(x,t),t) = x$ [1]. We thus have $\tilde{S}[\varphi] = S[\varphi(\xi(X,t),t)]$.

Proposition 1. *Extremizing $S[\phi]$ with respect to ϕ is equivalent to extremizing $\tilde{S}[\varphi]$ with respect to φ.*

Proof. The variational derivatives of S and $\tilde{S}$ are related by the formula

$$\delta\tilde{S}[\varphi] \cdot \delta\varphi(x,t) = \delta S[\varphi(\xi(X,t),t)] \cdot \delta\varphi(\xi(X,t),t). \tag{4}$$

Suppose $\phi(X,t)$ extremizes $S[\phi]$, i.e. $\delta S[\phi] \cdot \delta\phi = 0$ for all variations $\delta\phi$. Given the function $X(x,t)$, define $\varphi(x,t) = \phi(X(x,t),t)$. Then by the formula above we have $\delta\tilde{S}[\varphi] = 0$, so φ extremizes $\tilde{S}$.

[1] We allow a little abuse of notation here: X denotes both the argument of ξ and the change of variables $X(x,t)$. If we wanted to be more precise, we would write $X = h(x,t)$.

Conversely, suppose $\varphi(x,t)$ extremizes $\tilde{S}$, that is $\delta\tilde{S}[\varphi]\cdot\delta\varphi = 0$ for all variations $\delta\varphi$. Since we assume $X(.,t)$ is a homeomorphism, we can define $\phi(X,t) = \varphi(\xi(X,t),t)$. Note that an arbitrary variation $\delta\phi(X,t)$ induces the variation $\delta\varphi(x,t) = \delta\phi(X(x,t),t)$. Then we have $\delta S[\phi]\cdot\delta\phi = \delta\tilde{S}[\varphi]\cdot\delta\varphi = 0$ for all variations $\delta\phi$, so $\phi(X,t)$ extremizes $S[\phi]$.
□

The corresponding instantaneous Lagrangian $\tilde{L}: Q\times W\times\mathbb{R}\longrightarrow\mathbb{R}$ is

$$\tilde{L}[\varphi,\varphi_t,t] = \int_0^{X_{max}} \tilde{\mathcal{L}}(\varphi,\varphi_x,\varphi_t,t)\,dx \tag{5}$$

with the Lagrangian density

$$\tilde{\mathcal{L}}(\varphi,\varphi_x,\varphi_t,x,t) = \mathcal{L}\Big(\varphi,\frac{\varphi_x}{X_x},\varphi_t-\frac{\varphi_x X_t}{X_x}\Big)X_x. \tag{6}$$

The function spaces Q and W must be chosen appropriately for the problem at hand, so that (5) makes sense. For instance, for a free field we will have $Q = H^1([0,X_{max}])$ and $W = L^2([0,X_{max}])$. Since $X(x,t)$ is a function of t, we are looking at a time-dependent system. Even though the energy associated with (5) is not conserved, the energy of the original theory associated with (1)

$$E = \int_0^{X_{max}}\Big(\phi_t\frac{\partial\mathcal{L}}{\partial\phi_t}(\phi,\phi_X,\phi_t)-\mathcal{L}(\phi,\phi_X,\phi_t)\Big)\,dX \tag{7}$$

$$= \int_0^{X_{max}}\Big[\Big(\varphi_t-\frac{\varphi_x X_t}{X_x}\Big)\frac{\partial\mathcal{L}}{\partial\phi_t}\Big(\varphi,\frac{\varphi_x}{X_x},\varphi_t-\frac{\varphi_x X_t}{X_x}\Big)-\mathcal{L}\Big(\varphi,\frac{\varphi_x}{X_x},\varphi_t-\frac{\varphi_x X_t}{X_x}\Big)\Big]X_x\,dx \tag{8}$$

is conserved. To see this, note that if $\phi(X,t)$ extremizes $S[\phi]$ then $dE/dt = 0$ (computed from (7)). Trivially, this means that $dE/dt = 0$ when formula (8) is invoked as well. Moreover, as we have noted earlier, $\phi(X,t)$ extremizes $S[\phi]$ iff $\varphi(x,t)$ extremizes $\tilde{S}[\varphi]$. This means that the energy (8) is constant on solutions of the reparametrized theory.

2.2. *Spatial Finite Element discretization*

We begin with a discretization of the spatial dimension only, thus turning the original infinite-dimensional problem into a time-continuous finite-dimensional Lagrangian system. Let $\Delta x = X_{max}/(N+1)$ and define the reference uniform mesh $x_i = i\cdot\Delta x$ for $i = 0,1,...,N+1$, and the corresponding piecewise linear finite elements

$$\eta_i(x) = \begin{cases} \frac{x-x_{i-1}}{\Delta x}, & \text{if } x_{i-1}\le x\le x_i,\\ -\frac{x-x_{i+1}}{\Delta x}, & \text{if } x_i\le x\le x_{i+1},\\ 0, & \text{otherwise.}\end{cases} \tag{9}$$

We now restrict $X(x,t)$ to be of the form

$$X(x,t) = \sum_{i=0}^{N+1} X_i(t)\eta_i(x) \tag{10}$$

with $X_0(t) = 0$, $X_{N+1}(t) = X_{max}$ and arbitrary $X_i(t)$, $i = 1,2,...,N$ as long as $X(.,t)$ is a homeomorphism for all t. In our context of numerical computations, the functions $X_i(t)$ represent the current position of the i^{th} mesh point. Define the finite element spaces

$$Q_N = W_N = \text{span}(\eta_0,...,\eta_{N+1}) \tag{11}$$

and assume that $Q_N\subset Q$, $W_N\subset W$. Let us denote a generic element of Q_N by φ and a generic element of W_N by $\dot{\varphi}$. We have the decompositions

$$\varphi(x) = \sum_{i=0}^{N+1} y_i \eta_i(x), \qquad \dot{\varphi}(x) = \sum_{i=0}^{N+1} \dot{y}_i \eta_i(x). \tag{12}$$

The numbers $(y_i, \dot{y}_i)$ thus form natural (global) coordinates on $Q_N \times W_N$. We can now approximate the dynamics of system (5) in the finite-dimensional space $Q_N \times W_N$. Let us consider the restriction $\tilde{L}_N = \tilde{L}|_{Q_N \times W_N \times \mathbb{R}}$ of the Lagrangian (5) to $Q_N \times W_N \times \mathbb{R}$. In the chosen coordinates we have

$$\tilde{L}_N(y_0, ..., y_{N+1}, \dot{y}_0, ..., \dot{y}_{N+1}, t) = \tilde{L}\Big[\sum_{i=0}^{N+1} y_i \eta_i(x), \sum_{i=0}^{N+1} \dot{y}_i \eta_i(x), t\Big]. \tag{13}$$

Note that, given the boundary conditions (2), y_0, y_{N+1}, $\dot{y}_0$, and $\dot{y}_{N+1}$ are fixed. We will thus no longer write them as arguments of $\tilde{L}_N$.

The advantage of using a finite element discretization lies in the fact that the symplectic structure induced on $Q_N \times W_N$ by $\tilde{L}_N$ is strictly a restriction (i.e., a pull-back) of the (pre-)symplectic structure[2] on $Q \times W$. This establishes a direct link between symplectic integration of the finite-dimensional mechanical system $(Q_N \times W_N, \tilde{L}_N)$ and the infinite-dimensional field theory $(Q \times W, \tilde{L})$

2.3. DAE formulation and time integration

We now consider time integration of the Lagrangian system $(Q_N \times W_N, \tilde{L}_N)$. If the functions $X_i(t)$ are known, then one can perform variational integration in the standard way, that is, define the discrete Lagrangian $\tilde{L}_d : \mathbb{R} \times Q_N \times \mathbb{R} \times Q_N \to \mathbb{R}$ and solve the corresponding discrete Euler-Lagrange equations (see [2], [1]). Let $t_n = n \cdot \Delta t$ for $n = 0, 1, 2, \ldots$ be an increasing sequence of times and $\{y^0, y^1, \ldots\}$ the corresponding discrete path of the system in Q_N. The discrete Lagrangian L_d is an approximation to the exact discrete Lagrangian L_d^E, such that

$$\tilde{L}_d(t_n, y^n, t_{n+1}, y^{n+1}) \approx \tilde{L}_d^E(t_n, y^n, t_{n+1}, y^{n+1}) \equiv \int_{t_n}^{t_{n+1}} \tilde{L}_N(y(t), \dot{y}(t), t)\, dt, \tag{14}$$

where $y^n = (y_1^n, ..., y_N^n)$, $y^{n+1} = (y_1^{n+1}, ..., y_N^{n+1})$ and $y(t)$ is the solution of the Euler-Lagrange equations corresponding to $\tilde{L}_N$ with the boundary values $y(t_n) = y^n$, $y(t_{n+1}) = y^{n+1}$. Depending on the quadrature we use to approximate the integral in (14), we obtain different types of variational integrators. As will be discussed below, in r-adaptation one has to deal with stiff differential equations or differential-algebraic equations, therefore higher order implicit integration in time is advisable (see [11], [12]). We will employ variational partitioned Runge-Kutta methods. An s-stage Runge Kutta method is constructed by choosing

$$\tilde{L}_d(t_n, y^n, t_{n+1}, y^{n+1}) = (t_{n+1} - t_n) \sum_{i=1}^{s} b_i \tilde{L}_N(Y_i, \dot{Y}_i, t_i), \tag{15}$$

where $t_i = t_n + c_i(t_{n+1} - t_n)$, the right-hand side is extremized under the constraint $y^{n+1} = y^n + (t_{n+1} - t_n)\sum_{i=1}^{s} b_i \dot{Y}_i$, and the internal stage variables Y_i, $\dot{Y}_i$ are related by $Y_i = y^n + (t_{n+1} - t_n)\sum_{j=1}^{s} a_{ij} \dot{Y}_j$. It can be shown that the variational integrator with the discrete Lagrangian (15) is equivalent to an appropriately chosen symplectic partitioned Runge-Kutta method applied to the Hamiltonian system corresponding to $\tilde{L}_N$ (see [2], [1]). With this in mind we turn our semi-discrete Lagrangian system $(Q_N \times W_N, \tilde{L}_N)$ into the Hamiltonian system $(Q_N \times W_N^*, \tilde{H}_N)$ via the standard Legendre transform

[2] In most cases the symplectic structure of $(Q \times W, \tilde{L})$ is only weakly-nondegenerate; see [10]

$$\tilde{H}_N(y_1, ..., y_N, p_1, ..., p_N; X_1, ..., X_N, \dot{X}_1, ..., \dot{X}_N) = \sum_{i=1}^{N} p_i \dot{y}_i - \tilde{L}_N(y_1, ..., y_N, \dot{y}_1, ..., \dot{y}_N, t), \tag{16}$$

where $p_i = \partial \tilde{L}_N / \partial \dot{y}_i$ and we explicitly stated the dependence on the positions X_i and velocities $\dot{X}_i$ of the mesh points. The Hamiltonian equations take the form[3]

$$\begin{aligned} \dot{y}_i &= \frac{\partial \tilde{H}_N}{\partial p_i}\Big(y, p; X(t), \dot{X}(t)\Big), \\ \dot{p}_i &= -\frac{\partial \tilde{H}_N}{\partial y_i}\Big(y, p; X(t), \dot{X}(t)\Big). \end{aligned} \tag{17}$$

Suppose that the functions $X_i(t)$ are C^1 and H_N is smooth as a function of the y_i's, p_i's, X_i's and $\dot{X}_i$'s (note that these assumptions are used for simplicity, and can be easily relaxed if necessary, depending on the regularity of the considered Lagrangian system). Then the assumptions of Picard's theorem are satisfied and there exists a unique C^1 flow $F_{t_0,t} = (F^y_{t_0,t}, F^p_{t_0,t}) : Q_N \times W^*_N \to Q_N \times W^*_N$ for (17). This flow is symplectic.

However, in practice we do not know the X_i's and we in fact would like to be able to adjust them 'on the fly', based on the current behavior of the system. We will do that by introducing additional constraint functions $g_i(y_1, ..., y_N, X_1, ..., X_N)$ and demanding that the conditions $g_i = 0$ be satisfied at all times[4]. The choice of these functions will be discussed in Section 2.4. This leads to the following differential-algebraic system of index 1 (see [11], [12], [13])

$$\begin{aligned} \dot{y}_i &= \frac{\partial \tilde{H}_N}{\partial p_i}\Big(y, p; X, \dot{X}\Big), \\ \dot{p}_i &= -\frac{\partial \tilde{H}_N}{\partial y_i}\Big(y, p; X, \dot{X}\Big), \\ 0 &= g_i(y, X), \\ y_i(t_0) &= y_i^{(0)}, \\ p_i(t_0) &= p_i^{(0)} \end{aligned} \tag{18}$$

for $i = 1, ..., N$. Note that an initial condition for X is fixed by the constraints. This system is of index 1 because one has to differentiate the algebraic equations with respect to time once in order to reduce it to an implicit ODE system. In fact, the implicit system will take the form

---

[3] It is computationally more convenient to directly integrate the implicit Hamiltonian system $p_i = \partial \tilde{L}_N / \partial \dot{y}_i$, $\dot{p}_i = \partial \tilde{L}_N / \partial y_i$, but as long as system (1) is at least weakly-nondegenerate there is no theoretical issue with passing to the Hamiltonian formulation, which we do for the clarity of our exposition.

[4] In the context of Control Theory the constraints $g_i = 0$ are called *strict static state feedback*. See [9].

$$\begin{aligned}
\dot{y} &= \frac{\partial \tilde{H}_N}{\partial p}\Big(y, p; X, \dot{X}\Big), \\
\dot{p} &= -\frac{\partial \tilde{H}_N}{\partial y}\Big(y, p; X, \dot{X}\Big), \\
0 &= \frac{\partial g}{\partial y}(y, X)\dot{y} + \frac{\partial g}{\partial X}(y, X)\dot{X}, \\
y(t_0) &= y^{(0)}, \\
p(t_0) &= p^{(0)}, \\
X(t_0) &= X^{(0)},
\end{aligned} \tag{19}$$

where $X^{(0)}$ is a vector of arbitrary initial condition for the X_i's. Suppose again that H_N is a smooth function of y, p, X and $\dot{X}$. Futhermore, suppose that g is a C^1 function of y, X, and $\frac{\partial g}{\partial X} - \frac{\partial g}{\partial y}\frac{\partial^2 H_N}{\partial \dot{X} \partial p}$ is invertible with its inverse bounded in a neighborhood of the exact solution.[5] Then, by the Implicit Function Theorem equations (19) can be solved explicitly for $\dot{y}$, $\dot{p}$, $\dot{X}$ and the resulting explicit ODE system will satisfy the assumptions of Picard's theorem. Let $\big(y(t), p(t), X(t)\big)$ be the unique C^1 solution to this ODE system (and hence to (19)). We have the trivial result

Proposition 2. *If $g(y^{(0)}, X^{(0)}) = 0$, then $\big(y(t), p(t), X(t)\big)$ is a solution to (18).*[6]

In practice we would like to integrate system (18). A question arises in what sense is this system symplectic and in what sense a numerical integration scheme for this system can be regarded as variational. Let us address these issues.

Proposition 3. *Let $\big(y(t), p(t), X(t)\big)$ be a solution to (18) and use this $X(t)$ to form the Hamiltonian system (17). Then we have that*

$$y(t) = F^y_{t_0,t}(y^{(0)}, p^{(0)}), \qquad p(t) = F^p_{t_0,t}(y^{(0)}, p^{(0)})$$

and

$$g\Big(F^y_{t_0,t}(y^{(0)}, p^{(0)}), X(t)\Big) = 0,$$

where $F_{t_0,t}(\hat{y}, \hat{p})$ is the symplectic flow for (17).

Proof. Note that the first two equations of (18) are the same as (17), therefore $\big(y(t), p(t)\big)$ trivially satisfies (17) with the initial conditions $y(t_0) = y^{(0)}$ and $p(t_0) = p^{(0)}$. Since the flow map $F_{t_0,t}$ is unique, we must have $y(t) = F^y_{t_0,t}(y^{(0)}, p^{(0)})$ and $p(t) = F^p_{t_0,t}(y^{(0)}, p^{(0)})$. Then we also must have that $g\Big(F^y_{t_0,t}(y^{(0)}, p^{(0)}), X(t)\Big) = 0$, that is, the constraints are satisfied along one particular integral curve of (17) that passes through $(y^{(0)}, p^{(0)})$ at t_0.
□

Suppose we now would like to find a numerical approximation of the solution to (17) using an s-stage partitioned Runge-Kutta method with coefficients a_{ij}, b_i, $\bar{a}_{ij}$, $\bar{b}_i$, c_i ([14], [1]). The numerical scheme will take the form

---

[5] Again, these assumptions can be relaxed if necessary.

[6] Note that there might be other solutions, as for any given $y^{(0)}$ there might be more than one $X^{(0)}$ that solves the constraint equations.

$$
\begin{aligned}
\dot{Y}^i &= \frac{\partial \tilde{H}_N}{\partial p}\Big(Y^i, P^i; X(t_n + c_i\Delta t), \dot{X}(t_n + c_i\Delta t)\Big), \\
\dot{P}^i &= -\frac{\partial \tilde{H}_N}{\partial y}\Big(Y^i, P^i; X(t_n + c_i\Delta t), \dot{X}(t_n + c_i\Delta t)\Big), \\
Y^i &= y^n + \Delta t \sum_{j=1}^{s} a_{ij}\dot{Y}^j, \\
P^i &= p^n + \Delta t \sum_{j=1}^{s} \bar{a}_{ij}\dot{P}^j, \\
y^{n+1} &= y^n + \Delta t \sum_{i=1}^{s} b_i\dot{Y}^i, \\
p^{n+1} &= p^n + \Delta t \sum_{i=1}^{s} \bar{b}_i\dot{P}^i,
\end{aligned}
\tag{20}
$$

where Y^i, $\dot{Y}^i$, P^i, $\dot{P}^i$ are the internal stages and Δt is the integration timestep. Let us apply the same partitioned Runge-Kutta method to (18). In order to compute the internal stages Q^i, $\dot{Q}^i$ of the X variable we use the state-space form approach, that is, we demand that the constraints and their time derivatives be satisfied (see [12]). The new step value X^{n+1} is computed by solving the constraints as well. The resulting numerical scheme is thus

$$
\begin{aligned}
\dot{Y}^i &= \frac{\partial \tilde{H}_N}{\partial p}\Big(Y^i, P^i; Q^i, \dot{Q}^i\Big), \\
\dot{P}^i &= -\frac{\partial \tilde{H}_N}{\partial y}\Big(Y^i, P^i; Q^i, \dot{Q}^i\Big), \\
Y^i &= y^n + \Delta t \sum_{j=1}^{s} a_{ij}\dot{Y}^j, \\
P^i &= p^n + \Delta t \sum_{j=1}^{s} \bar{a}_{ij}\dot{P}^j, \\
0 &= g(Y^i, Q^i), \\
0 &= \frac{\partial g}{\partial y}(Y^i, Q^i)\,\dot{Y}^i + \frac{\partial g}{\partial X}(Y^i, Q^i)\,\dot{Q}^i, \\
y^{n+1} &= y^n + \Delta t \sum_{i=1}^{s} b_i\dot{Y}^i, \\
p^{n+1} &= p^n + \Delta t \sum_{i=1}^{s} \bar{b}_i\dot{P}^i, \\
0 &= g(y^{n+1}, X^{n+1}).
\end{aligned}
\tag{21}
$$

We have the following trivial observation.

Proposition 4. *If $X(t)$ is defined to be a C^1 interpolation of the internal stages Q^i, $\dot{Q}^i$ at times $t_n + c_i\Delta t$ (that is, if the values $X(t_n + c_i\Delta t)$, $\dot{X}(t_n + c_i\Delta t)$ coincide with Q^i, $\dot{Q}^i$), then the schemes (20) and (21) give the same numerical approximations y^n, p^n to the exact solution $y(t)$, $p(t)$.*

Intuitively, Proposition 4 states that we can apply a symplectic partitioned Runge-Kutta method to the DAE system (18), which solves both for $X(t)$ and $(y(t), p(t))$, and the result will be the same as if we performed a symplectic integration of the Hamiltonian system (17) for $(y(t), p(t))$ with a *known* $X(t)$.

2.4. Moving mesh partial differential equations

The concept of equidistribution is the most popular paradigm of r-adaptation (see [7], [8]). Given a continuous mesh density function $\rho(X)$, the equidistribution principle seeks to find a mesh $0 = X_0 < X_1 < ... < X_{N+1} = X_{max}$ such that the following holds

$$\int_0^{X_1} \rho(X)\,dX = \int_{X_1}^{X_2} \rho(X)\,dX = ... = \int_{X_N}^{X_{max}} \rho(X)\,dX, \tag{22}$$

that is, the quantity represented by the density function is equidistributed among all cells. In the continuous setting we will say that the reparametrization $X = X(x)$ equidistributes $\rho(X)$ if

$$\int_0^{X(x)} \rho(X)\,dX = \frac{x}{X_{max}}\sigma, \tag{23}$$

where $\sigma = \int_0^{X_{max}} \rho(X)\,dX$ is the total amount of the equidistributed quantity. Differentiate this equation with respect to x to obtain

$$\rho(X(x))\frac{\partial X}{\partial x} = \frac{1}{X_{max}}\sigma. \tag{24}$$

It is still a global condition in the sense that σ has to be known. For computational purposes it is convenient to differentiate this relation again and consider the following partial differential equation

$$\frac{\partial}{\partial x}\Big(\rho(X(x))\frac{\partial X}{\partial x}\Big) = 0 \tag{25}$$

with the boundary conditions $X(0) = 0$, $X(X_{max}) = X_{max}$. The choice of the mesh density function $\rho(X)$ is typically problem-dependent and the subject of much research. A popular example is the generalized solution arclength given by

$$\rho = \sqrt{1 + \alpha^2\Big(\frac{\partial\phi}{\partial X}\Big)^2} = \sqrt{1 + \alpha^2\Big(\frac{\varphi_x}{X_x}\Big)^2}. \tag{26}$$

It is often used to construct meshes that can follow moving fronts with locally high gradients ([7], [8]). With this choice, equation (25) is equivalent to

$$\alpha^2\varphi_x\varphi_{xx} + X_x X_{xx} = 0, \tag{27}$$

assuming $X_x > 0$, which we demand anyway. A finite difference discretization on the mesh $x_i = i \cdot \Delta x$ gives us the set of contraints

$$\begin{aligned} g_i(y_1, ..., y_N, X_1, ..., X_N) = \\ \alpha^2(y_{i+1} - y_i)^2 + (X_{i+1} - X_i)^2 - \alpha^2(y_i - y_{i-1})^2 - (X_i - X_{i-1})^2 = 0, \end{aligned} \tag{28}$$

with the previously defined y_i's and X_i's. This set of constraints can be used in (18).

2.5. Example

To illustrate these ideas let us consider the Lagrangian density

$$\mathcal{L}(\phi, \phi_X, \phi_t) = \frac{1}{2}\phi_t^2 - W(\phi_X). \tag{29}$$

The reparametrized Lagrangian (5) takes the form

$$\tilde{L}[\varphi,\varphi_t,t]=\int_0^{X_{max}}\left[\frac{1}{2}X_x\Big(\varphi_t-\frac{\varphi_x}{X_x}X_t\Big)^2-W\Big(\frac{\varphi_x}{X_x}\Big)X_x\right]dx. \tag{30}$$

Let $N=1$ and $\phi_L=\phi_R=0$. Then

$$\varphi(x,t)=y_1(t)\eta_1(x),\qquad X(x,t)=X_1(t)\eta_1(x)+X_{max}\,\eta_2(x). \tag{31}$$

The semi-discrete Lagrangian is

$$\begin{aligned}\tilde{L}_N(y_1,\dot{y}_1,t)=&\frac{X_1(t)}{6}\Big(\dot{y}_1-\frac{y_1}{X_1(t)}\dot{X}_1(t)\Big)^2+\frac{X_{max}-X_1(t)}{6}\Big(\dot{y}_1+\frac{y_1}{X_{max}-X_1(t)}\dot{X}_1(t)\Big)^2\\&-W\Big(\frac{y_1}{X_1(t)}\Big)X_1(t)-W\Big(-\frac{y_1}{X_{max}-X_1(t)}\Big)\big(X_{max}-X_1(t)\big).\end{aligned} \tag{32}$$

The Legendre transform gives $p_1=\partial\tilde{L}_N/\partial\dot{y}_1=X_{max}\dot{y}_1/3$, hence the semi-discrete Hamiltonian is

$$\begin{aligned}\tilde{H}_N(y_1,p_1;X_1,\dot{X}_1)=&\frac{3}{2X_{max}}p_1^2-\frac{1}{6}\frac{X_{max}\dot{X}_1^2}{X_1(X_{max}-X_1)}y_1^2\\&+W\Big(\frac{y_1}{X_1}\Big)X_1+W\Big(-\frac{y_1}{X_{max}-X_1}\Big)(X_{max}-X_1).\end{aligned} \tag{33}$$

The corresponding DAE system is

$$\begin{aligned}\dot{y}_1&=\frac{3}{X_{max}}p_1,\\\dot{p}_1&=\frac{1}{3}\frac{X_{max}\dot{X}_1^2}{X_1(X_{max}-X_1)}y_1-W'\Big(\frac{y_1}{X_1}\Big)+W'\Big(-\frac{y_1}{X_{max}-X_1}\Big),\\0&=g_1(y_1,X_1).\end{aligned} \tag{34}$$

This system is to be solved for the unknown functions $y_1(t)$, $p_1(t)$ and $X_1(t)$. It is of index 1, because we have three unknown functions and only two differential equations — the algebraic equation has to be differentiated once in order to obtain a missing ODE.

2.6. Backward error analysis

The true power of symplectic integration of Hamiltonian equations is revealed through backward error analysis: it can be shown that a symplectic integrator for a Hamiltonian system with the Hamiltonian $H(q,p)$ defines the *exact* flow for a nearby Hamiltonian system, whose Hamiltonian can be expressed as the asymptotic series

$$\mathscr{H}(q,p)=H(q,p)+\Delta t H_2(q,p)+\Delta t^2 H_3(q,p)+\ldots \tag{35}$$

Owing to this fact, under some additional assumptions symplectic numerical schemes nearly conserve the original Hamiltonian $H(q,p)$ over exponentially long time intervals. See [1] for details.

Let us briefly review the results of backward error analysis for the integrator (21). Suppose $g(y,X)$ satisfies the assumptions of the Implicit Function Theorem. Then, at least locally, we can solve the constraint $X=h(y)$. The Hamiltonian DAE system (18) can be then written as the following (implicit) ODE system for y and p

$$\dot{y} = \frac{\partial \tilde{H}_N}{\partial p}\Big(y, p; h(y), h'(y)\dot{y}\Big),$$
$$\dot{p} = -\frac{\partial \tilde{H}_N}{\partial y}\Big(y, p; h(y), h'(y)\dot{y}\Big). \tag{36}$$

Since we used the state-space formulation, the numerical scheme (21) is equivalent to applying the same partitioned Runge-Kutta method to (36), that is, we have $Q^i = h(Y^i)$ and $\dot{Q}^i = h'(Y^i)\dot{Y}^i$. We computed the corresponding modified equation for several symplectic methods, namely Gauss and Lobatto IIIA-IIIB quadratures. Unfortunately, none of the quadratures resulted in a form akin to (36) for some modified Hamiltonian function $\tilde{\mathscr{H}}_N$ related to $\tilde{H}_N$ by a series similar to (35). This hints at the fact that we should not expect this integrator to show excellent energy conservation over long integration times. One could also consider the implicit ODE system (19), which has an obvious triple partitioned structure, and apply a different Runge-Kutta method to each variable y, p and X. Although we did not pursue this idea further, it seems unlikely it would bring a desirable result.

We therefore conclude that the control-theoretic strategy, while yielding a perfectly legitimate numerical method, does not take the full advantage of the underlying geometric structures. Let us point out that, while we used a variational discretization of the governing physical PDE, the mesh equations were coupled in a manner that is typical of the existing r-adaptive methods (see [7], [8]). We now turn our attention to a second approach, which offers a novel way of coupling the mesh equations to the physical equations.

3. Lagrange multiplier approach to r-adaptation

As we saw in Section 2, discretization of the variational principle alone is not sufficient if we would like to accurately capture the geometric properties of the physical system described by (1). In this section we propose a new technique of coupling the mesh equations to the physical equations. Our idea is based on the observation that in r-adaptation the number of mesh points is constant, therefore we can treat them as pseudo-particles, and we can incorporate their dynamics into the variational principle. We show that this strategy results in integrators that much better preserve the energy of the considered system.

3.1. Reparametrized Lagrangian

In this approach, we treat $X(x,t)$ as an independent field, that is, another degree of freedom, and we will treat the 'modified' action (3) as a functional of both φ and X: $\tilde{S} = \tilde{S}[\varphi, X]$. For the purpose of the derivations below, we assume that $\varphi(.,t)$ and $X(.,t)$ are continuous and piecewise C^1. One could consider the closure of this space in the topology of either Hilbert or Banach space of sufficiently integrable functions and interpret differentiation in a sufficiently weak sense, but this functional-analytic aspect is of little importance for the developments in this section. We refer the interested reader to [15] and [16]. As in Section 2.1, let $\xi(X,t)$ be the function such that $\xi(.,t) = X(.,t)^{-1}$, that is $\xi(X(x,t),t) = x$. Then $\tilde{S}[\varphi, X] = S[\varphi(\xi(X,t),t)]$. We begin with two propositions and one corollary which will be important for the rest of our exposition.

Proposition 5. *Extremizing $S[\phi]$ with respect to ϕ is equivalent to extremizing $\tilde{S}[\varphi, X]$ with respect to both φ and X.*

Proof. The variational derivatives of S and $\tilde{S}$ are related by the formula

$$\delta_1 \tilde{S}[\varphi, X] \cdot \delta\varphi(x,t) = \delta S[\varphi(\xi(X,t),t)] \cdot \delta\varphi(\xi(X,t),t),$$
$$\delta_2 \tilde{S}[\varphi, X] \cdot \delta X(x,t) = \delta S[\varphi(\xi(X,t),t)] \cdot \Big(- \frac{\varphi_x(\xi(X,t),t)}{X_x(\xi(X,t),t)} \delta X(\xi(X,t),t) \Big), \tag{37}$$

where δ_1 and δ_2 denote differentiation with respect to the first and second argument, respectively. Suppose $\phi(X,t)$ extremizes $S[\phi]$, i.e. $\delta S[\phi] \cdot \delta\phi = 0$ for all variations $\delta\phi$. Choose an arbitrary $X(x,t)$, such that $X(.,t)$ is a (sufficiently smooth) homeomorphism and define $\varphi(x,t) = \phi(X(x,t),t)$. Then by the formula above we have $\delta_1 \tilde{S}[\varphi, X] = 0$ and $\delta_2 \tilde{S}[\varphi, X] = 0$, so the pair (φ, X) extremizes $\tilde{S}$. Conversely, suppose the pair (φ, X) extremizes $\tilde{S}$, that is $\delta_1 \tilde{S}[\varphi, X] \cdot \delta\varphi = 0$ and $\delta_2 \tilde{S}[\varphi, X] \cdot \delta X = 0$ for all variations $\delta\varphi$ and δX. Since we assume $X(.,t)$ is a homeomorphism, we can define $\phi(X,t) = \varphi(\xi(X,t),t)$. Note that an arbitrary variation $\delta\phi(X,t)$ induces the variation $\delta\varphi(x,t) = \delta\phi(X(x,t),t)$. Then we have $\delta S[\phi] \cdot \delta\phi = \delta_1 \tilde{S}[\varphi, X] \cdot \delta\varphi = 0$ for all variations $\delta\phi$, so $\phi(X,t)$ extremizes $S[\phi]$.
□

Proposition 6. *The equation $\delta_2 \tilde{S}[\varphi, X] = 0$ is implied by the equation $\delta_1 \tilde{S}[\varphi, X] = 0$.*

Proof. As we saw in the proof of Proposition 5, the condition $\delta_1 \tilde{S}[\varphi, X] \cdot \delta\varphi = 0$ implies $\delta S = 0$. By (37), this in turn implies $\delta_2 \tilde{S}[\varphi, X] \cdot \delta X = 0$ for all δX. Note that this argument cannot be reversed: $\delta_2 \tilde{S}[\varphi, X] \cdot \delta X = 0$ does not imply $\delta S = 0$ when $\varphi_x = 0$.
□

Corollary 1. *The field theory described by $\tilde{S}[\varphi, X]$ is degenerate and the solutions to the Euler-Lagrange equations are not unique.*

3.2. Spatial Finite Element discretization

The Lagrangian of the 'reparametrized' theory $\tilde{L} : Q \times G \times W \times Z \longrightarrow \mathbb{R}$,

$$\tilde{L}[\varphi, X, \varphi_t, X_t] = \int_0^{X_{max}} \mathcal{L}\Big(\varphi, \frac{\varphi_x}{X_x}, \varphi_t - \frac{\varphi_x X_t}{X_x}\Big) X_x \, dx, \tag{38}$$

has the same form as (5) (we only treat it as a functional of X and X_t as well), where Q, G, W and Z are spaces of continuous and piecewise C^1 functions, as mentioned before. We again let $\Delta x = X_{max}/(N+1)$ and define the uniform mesh $x_i = i \cdot \Delta x$ for $i = 0, 1, ..., N+1$. Define the finite element spaces

$$Q_N = G_N = W_N = Z_N = \text{span}(\eta_0, ..., \eta_{N+1}), \tag{39}$$

where we used the finite elements (9). We have $Q_N \subset Q$, $G_N \subset G$, $W_N \subset W$, $Z_N \subset Z$. In addition to (12) we also consider

$$X(x) = \sum_{i=0}^{N+1} X_i \eta_i(x), \qquad \dot{X}(x) = \sum_{i=0}^{N+1} \dot{X}_i \eta_i(x). \tag{40}$$

The numbers $(y_i, X_i, \dot{y}_i, \dot{X}_i)$ thus form natural (global) coordinates on $Q_N \times G_N \times W_N \times Z_N$. We again consider the restricted Lagrangian $\tilde{L}_N = \tilde{L}|_{Q_N \times G_N \times W_N \times Z_N}$. In the chosen coordinates

$$\tilde{L}_N(y_1, ..., y_N, X_1, ..., X_N, \dot{y}_1, ..., \dot{y}_N, \dot{X}_1, ..., \dot{X}_N) = \tilde{L}\Big[\varphi(x), X(x), \dot{\varphi}(x), \dot{X}(x)\Big], \tag{41}$$

where $\varphi(x)$, $X(x)$, $\dot{\varphi}(x)$, $\dot{X}(x)$ are defined by (12) and (40). Once again, we refrain from writing y_0, y_{N+1}, $\dot{y}_0$, $\dot{y}_{N+1}$, X_0, X_{N+1}, $\dot{X}_0$ and $\dot{X}_{N+1}$ as arguments of $\tilde{L}_N$ in the remainder of this section, as those are not actual degrees of freedom.

3.3. Invertibility of the Legendre Transform

For simplicity, let us restrict our considerations to Lagrangian densities of the form

$$\mathcal{L}(\phi, \phi_X, \phi_t) = \frac{1}{2}\phi_t^2 - R(\phi_X, \phi). \tag{42}$$

We chose a kinetic term that is most common in applications. The corresponding 'reparametrized' Lagrangian is

$$\tilde{L}[\varphi, X, \varphi_t, X_t] = \int_0^{X_{max}} \frac{1}{2} X_x \Big(\varphi_t - \frac{\varphi_x}{X_x} X_t\Big)^2 dx - \dots, \tag{43}$$

where we kept only the terms that involve the velocities φ_t and X_t. The semi-discrete Lagrangian becomes

$$\begin{aligned}\tilde{L}_N = \sum_{i=0}^{N} \frac{X_{i+1}-X_i}{6} \Big[\Big(\dot{y}_i - \frac{y_{i+1}-y_i}{X_{i+1}-X_i}\dot{X}_i\Big)^2 + \Big(\dot{y}_i - \frac{y_{i+1}-y_i}{X_{i+1}-X_i}\dot{X}_i\Big)\Big(\dot{y}_{i+1} - \frac{y_{i+1}-y_i}{X_{i+1}-X_i}\dot{X}_{i+1}\Big) \\ + \Big(\dot{y}_{i+1} - \frac{y_{i+1}-y_i}{X_{i+1}-X_i}\dot{X}_{i+1}\Big)^2\Big] - \dots \end{aligned} \tag{44}$$

Let us define the conjugate momenta via the Legendre Transform

$$p_i = \frac{\partial \tilde{L}_N}{\partial \dot{y}_i}, \qquad S_i = \frac{\partial \tilde{L}_N}{\partial \dot{X}_i}, \qquad i = 1, 2, ..., N. \tag{45}$$

This can be written as

$$\begin{pmatrix} p_1 \\ S_1 \\ \vdots \\ p_N \\ S_N \end{pmatrix} = \tilde{M}_N(y, X) \cdot \begin{pmatrix} \dot{y}_1 \\ \dot{X}_1 \\ \vdots \\ \dot{y}_N \\ \dot{X}_N \end{pmatrix}, \tag{46}$$

where the $2N \times 2N$ mass matrix $\tilde{M}_N(y, X)$ has the following block tridiagonal structure

$$\tilde{M}_N(y, X) = \begin{pmatrix} A_1 & B_1 & & & & \\ B_1 & A_2 & B_2 & & & \\ & B_2 & A_3 & B_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & B_{N-1} \\ & & & & B_{N-1} & A_N \end{pmatrix}, \tag{47}$$

with the 2×2 blocks

$$A_i = \begin{pmatrix} \frac{1}{3}\delta_{i-1} + \frac{1}{3}\delta_i & -\frac{1}{3}\delta_{i-1}\gamma_{i-1} - \frac{1}{3}\delta_i\gamma_i \\ -\frac{1}{3}\delta_{i-1}\gamma_{i-1} - \frac{1}{3}\delta_i\gamma_i & \frac{1}{3}\delta_{i-1}\gamma_{i-1}^2 + \frac{1}{3}\delta_i\gamma_i^2 \end{pmatrix}, \qquad B_i = \begin{pmatrix} \frac{1}{6}\delta_i & -\frac{1}{6}\delta_i\gamma_i \\ -\frac{1}{6}\delta_i\gamma_i & \frac{1}{6}\delta_i\gamma_i^2 \end{pmatrix}, \tag{48}$$

where

$$\delta_i = X_{i+1} - X_i, \qquad \gamma_i = \frac{y_{i+1}-y_i}{X_{i+1}-X_i}. \tag{49}$$

From now on we will always assume $\delta_i > 0$, as we demand that $X(x) = \sum_{i=0}^{N+1} X_i \eta_i(x)$ be a homeomorphism. We also have

$$\det A_i = \frac{1}{9}\delta_{i-1}\delta_i(\gamma_{i-1} - \gamma_i)^2. \tag{50}$$

Proposition 7. *The mass matrix $\tilde{M}_N(y, X)$ is non-singular almost everywhere (as a function of the y_i's and X_i's) and singular iff $\gamma_{i-1} = \gamma_i$ for some i.*

Proof. We will compute the determinant of $\tilde{M}_N(y, X)$ by transforming (47) into a block upper triangular form by zeroing the blocks B_i below the diagonal. Let us start with the block B_1. We use linear combinations of the first two rows of the mass matrix to zero the elements of the block B_1 below the diagonal. Suppose $\gamma_0 = \gamma_1$. Then it is easy to see that the first two rows of the mass matrix are not linearly independent, so the determinant of the mass matrix is zero. Assume $\gamma_0 \neq \gamma_1$. Then by (50) the block A_1 is invertible. We multiply the first two rows of the mass matrix by $B_1 A_1^{-1}$ and subtract the result from the third and fourth rows. This zeroes the block B_1 below the diagonal and replaces the block A_2 by

$$C_2 = A_2 - B_1 A_1^{-1} B_1. \tag{51}$$

We now zero the block B_2 below the diagonal in a similar fashion. After $n-1$ steps of this procedure the mass matrix is transformed into

$$\begin{pmatrix} C_1 & B_1 & & & & & \\ & C_2 & B_2 & & & & \\ & & \ddots & \ddots & & & \\ & & & C_n & B_n & & \\ & & & B_n & A_{n+1} & \ddots & \\ & & & & \ddots & \ddots & B_{N-1} \\ & & & & & B_{N-1} & A_N \end{pmatrix}. \tag{52}$$

In a moment we will see that C_n is singular iff $\gamma_{n-1} = \gamma_n$ and in that case the two rows of the matrix above that contain C_n and B_n are linearly dependent, thus making the mass matrix singular. Suppose $\gamma_{n-1} \neq \gamma_n$, so that C_n is invertible. In the next step of our procedure the block A_{n+1} is replaced by

$$C_{n+1} = A_{n+1} - B_n C_n^{-1} B_n. \tag{53}$$

Together with the condition $C_1 = A_1$ this gives us a recurrence. By induction on n we find that

$$C_n = \begin{pmatrix} \frac{1}{4}\delta_{n-1} + \frac{1}{3}\delta_n & -\frac{1}{4}\delta_{n-1}\gamma_{n-1} - \frac{1}{3}\delta_n\gamma_n \\ -\frac{1}{4}\delta_{n-1}\gamma_{n-1} - \frac{1}{3}\delta_n\gamma_n & \frac{1}{4}\delta_{n-1}\gamma_{n-1}^2 + \frac{1}{3}\delta_n\gamma_n^2 \end{pmatrix} \tag{54}$$

and

$$\det C_i = \frac{1}{12}\delta_{i-1}\delta_i(\gamma_{i-1} - \gamma_i)^2, \tag{55}$$

which justifies our assumptions on the invertibility of the blocks C_i. We can now express the determinant of the mass matrix as $\det C_1 \cdot \ldots \cdot \det C_N$. The final formula is

$$\det \tilde{M}_N(y, X) = \frac{\delta_0\delta_1^2 \ldots \delta_{N-1}^2\delta_N}{9 \cdot 12^{N-1}}(\gamma_0 - \gamma_1)^2 \ldots (\gamma_{N-1} - \gamma_N)^2. \tag{56}$$

We see that the mass matrix becomes singular iff $\gamma_{i-1} = \gamma_i$ for some i and this condition defines a measure zero subset of $\mathbb{R}^{2N}$.

□

Remark I.

This result shows that the finite-dimensional system described by the semi-discrete Lagrangian (44) is non-degenerate almost everywhere. This means that, unlike in the continuous case, the Euler-Lagrange equations corresponding to the variations of the y_i's and X_i's are independent of each other (almost everywhere) and the equations corresponding to the X_i's are in fact necessary for the correct description of the dynamics. This can also be seen in a more general way. Owing to the fact we are considering a finite element approximation, the semi-discrete action functional $\tilde{S}_N$ is simply a restriction of $\tilde{S}$, and therefore formulas (37) still hold. The corresponding Euler-Lagrange equations take the form

$$\begin{aligned} \delta_1 \tilde{S}[\varphi, X] \cdot \delta\varphi(x,t) &= 0, \\ \delta_2 \tilde{S}[\varphi, X] \cdot \delta X(x,t) &= 0, \end{aligned} \tag{57}$$

which must hold for all variations $\delta\varphi(x,t) = \sum_{i=1}^N \delta y_i(t)\eta_i(x)$ and $\delta X(x,t) = \sum_{i=1}^N \delta X_i(t)\eta_i(x)$. Since we are working in a finite dimensional subspace, the second equation now does not follow from the first equation. To see this, consider a particular variation $\delta X(x,t) = \delta X_k(t)\eta_k(x)$ for some k, where $\delta X_k \not\equiv 0$. Then we have

$$-\frac{\varphi_x}{X_x}\delta X_k(t) = \begin{cases} -\gamma_{k-1}\,\delta X_k(t)\,\eta_k(x), & \text{if } x_{k-1} \le x \le x_k, \\ -\gamma_k\,\delta X_k(t)\,\eta_k(x), & \text{if } x_k \le x \le x_{k+1}, \\ 0, & \text{otherwise}, \end{cases} \tag{58}$$

which is discontinuous at $x = x_k$ and cannot be expressed as $\sum_{i=1}^N \delta y_i(t)\eta_i(x)$ for any $\delta y_i(t)$, unless $\gamma_{k-1} = \gamma_k$. Therefore, we cannot invoke the first equation to show that $\delta_2 \tilde{S}[\varphi, X] \cdot \delta X(x,t) = 0$. The second equation becomes independent.

Remark II.

It is also instructive to realize what exactly happens when $\gamma_{k-1} = \gamma_k$. This means that locally in the interval $[X_{k-1}, X_{k+1}]$ the field $\phi(X,t)$ is a straight line with the slope γ_k. It also means that there are infinitely many values (X_k, y_k) that reproduce the same local shape of $\phi(X,t)$. This reflects the arbitrariness of $X(x,t)$ in the infinite-dimensional setting. In the finite element setting, however, this holds only when the points (X_{k-1}, y_{k-1}), (X_k, y_k) and (X_{k+1}, y_{k+1}) line up. Otherwise any change to the middle point changes the shape of $\phi(X,t)$. See Figure 1.

3.4. Existence and uniqueness of solutions

Since the Legendre Transform (46) becomes singular at some points, this raises a question about the existence and uniqueness of the solutions to the Euler-Lagrange equations (57). In this section we provide a partial answer to this problem. We will begin by computing the Lagrangian symplectic form

$$\tilde{\Omega}_N = \sum_{i=1}^N dy_i \wedge dp_i + dX_i \wedge dS_i, \tag{59}$$

where p_i and S_i are given by (45). For notational convenience we will collectively denote $q = (y_1, X_1, ..., y_N, X_N)^T$ and $\dot{q} = (\dot{y}_1, \dot{X}_1, ..., \dot{y}_N, \dot{X}_N)^T$. Then in the ordered basis $(\frac{\partial}{\partial q_1}, ..., \frac{\partial}{\partial q_{2N}}, \frac{\partial}{\partial \dot{q}_1}, ..., \frac{\partial}{\partial \dot{q}_{2N}})$ the symplectic form can be represented by the matrix

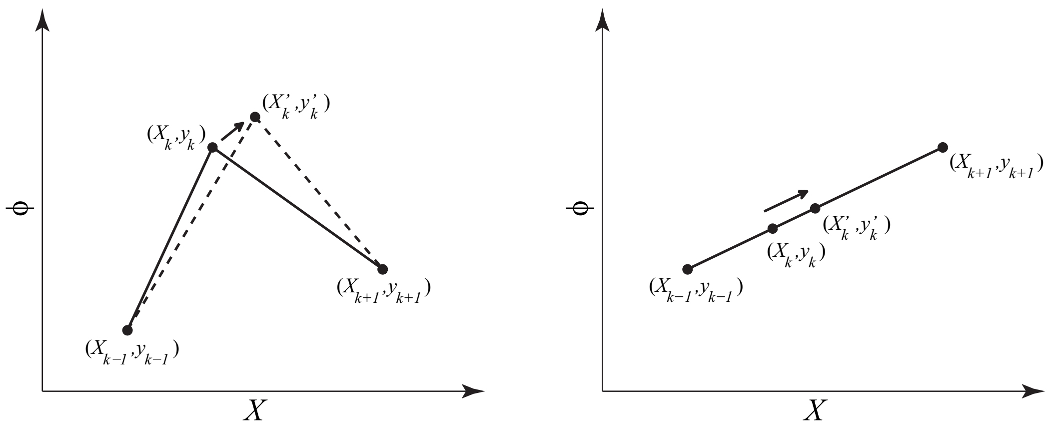

Figure 1. *Left:* If $\gamma_{k-1} \neq \gamma_k$, then any change to the middle point changes the local shape of $\phi(X,t)$. *Right:* If $\gamma_{k-1} = \gamma_k$, then there are infinitely many possible positions for (X_k, y_k) that reproduce the local linear shape of $\phi(X,t)$.

$$\tilde{\Omega}_N(q,\dot{q}) = \begin{pmatrix} \tilde{\Delta}_N(q,\dot{q}) & \tilde{M}_N(q) \\ -\tilde{M}_N(q) & 0 \end{pmatrix}, \tag{60}$$

where the $2N \times 2N$ block $\tilde{\Delta}_N(q,\dot{q})$ has the further block tridiagonal structure

$$\tilde{\Delta}_N(q,\dot{q}) = \begin{pmatrix} \Gamma_1 & \Lambda_1 & & & & \\ -\Lambda_1^T & \Gamma_2 & \Lambda_2 & & & \\ & -\Lambda_2^T & \Gamma_3 & \Lambda_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & \Lambda_{N-1} \\ & & & & -\Lambda_{N-1}^T & \Gamma_N \end{pmatrix} \tag{61}$$

with the 2×2 blocks

$$\begin{aligned} \Gamma_i &= \begin{pmatrix} 0 & -\frac{\dot{y}_{i+1}-\dot{y}_{i-1}}{3} - \frac{\dot{X}_{i-1}+2\dot{X}_i}{3}\gamma_{i-1} + \frac{2\dot{X}_i+\dot{X}_{i+1}}{3}\gamma_i \\ \frac{\dot{y}_{i+1}-\dot{y}_{i-1}}{3} + \frac{\dot{X}_{i-1}+2\dot{X}_i}{3}\gamma_{i-1} - \frac{2\dot{X}_i+\dot{X}_{i+1}}{3}\gamma_i & 0 \end{pmatrix}, \\ \Lambda_i &= \begin{pmatrix} -\frac{\dot{X}_i+\dot{X}_{i+1}}{2} & -\frac{\dot{y}_{i+1}-\dot{y}_i}{6} + \frac{\dot{X}_i+2\dot{X}_{i+1}}{3}\gamma_i \\ \frac{\dot{y}_{i+1}-\dot{y}_i}{6} + \frac{2\dot{X}_i+\dot{X}_{i+1}}{3}\gamma_i & -\frac{\dot{X}_i+\dot{X}_{i+1}}{2}\gamma_i^2 \end{pmatrix}. \end{aligned} \tag{62}$$

In this form, it is easy to see that

$$\det \tilde{\Omega}_N(q,\dot{q}) = \Big(\det \tilde{M}_N(q)\Big)^2, \tag{63}$$

so the symplectic form is singular whenever the mass matrix is.

The energy corresponding to the Lagrangian (44) can be written as

$$\tilde{E}_N(q,\dot{q}) = \frac{1}{2}\dot{q}^T \tilde{M}_N(q)\,\dot{q} + \sum_{k=0}^{N} \int_{x_k}^{x_{k+1}} R\Big(\gamma_k, y_k\eta_k(x) + y_{k+1}\eta_{k+1}(x)\Big)\frac{X_{k+1}-X_k}{\Delta x}\,dx. \tag{64}$$

In the chosen coordinates, $d\tilde{E}_N$ can be represented by the row vector $d\tilde{E}_N = (\partial\tilde{E}_N/\partial q_1, ..., \partial\tilde{E}_N/\partial\dot{q}_{2N})$. It turns out that

$$d\tilde{E}_N^T(q,\dot{q}) = \begin{pmatrix} \xi \\ \tilde{M}_N(q)\dot{q} \end{pmatrix}, \tag{65}$$

where the vector ξ has the following block structure

$$\xi = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_N \end{pmatrix}. \tag{66}$$

Each of these blocks has the form $\xi_k = (\xi_{k,1}, \xi_{k,2})^T$. Through basic algebraic manipulations and integration by parts, one finds that

$$\begin{aligned}
\xi_{k,1} = \; & \frac{\dot{y}_{k+1}(2\dot{X}_{k+1}+\dot{X}_k) + \dot{y}_k(\dot{X}_{k+1}-\dot{X}_{k-1}) - \dot{y}_{k-1}(\dot{X}_k + 2\dot{X}_{k-1})}{6} \\
& + \frac{\dot{X}_k^2 + \dot{X}_k\dot{X}_{k-1} + \dot{X}_{k-1}^2}{3}\gamma_{k-1} - \frac{\dot{X}_{k+1}^2 + \dot{X}_{k+1}\dot{X}_k + \dot{X}_k^2}{3}\gamma_k \\
& + \frac{1}{\Delta x}\int_{x_{k-1}}^{x_k} \frac{\partial R}{\partial \phi_X}\Big(\gamma_{k-1}, y_{k-1}\eta_{k-1}(x) + y_k\eta_k(x)\Big)\,dx \\
& - \frac{1}{\Delta x}\int_{x_k}^{x_{k+1}} \frac{\partial R}{\partial \phi_X}\Big(\gamma_k, y_k\eta_k(x) + y_{k+1}\eta_{k+1}(x)\Big)\,dx \\
& + \frac{1}{\gamma_{k-1}}\Big[R(\gamma_{k-1}, y_k) - \frac{1}{\Delta x}\int_{x_{k-1}}^{x_k} R\Big(\gamma_{k-1}, y_{k-1}\eta_{k-1}(x) + y_k\eta_k(x)\Big)\,dx\Big] \\
& - \frac{1}{\gamma_k}\Big[R(\gamma_k, y_k) - \frac{1}{\Delta x}\int_{x_k}^{x_{k+1}} R\Big(\gamma_k, y_k\eta_k(x) + y_{k+1}\eta_{k+1}(x)\Big)\,dx\Big],
\end{aligned} \tag{67}$$

and

$$\begin{aligned}
\xi_{k,2} = \; & \frac{\dot{y}_{k-1}^2 + \dot{y}_{k-1}\dot{y}_k - \dot{y}_k\dot{y}_{k+1} - \dot{y}_{k+1}^2}{6} \\
& - \frac{\dot{X}_k^2 + \dot{X}_k\dot{X}_{k-1} + \dot{X}_{k-1}^2}{6}\gamma_{k-1}^2 + \frac{\dot{X}_{k+1}^2 + \dot{X}_{k+1}\dot{X}_k + \dot{X}_k^2}{6}\gamma_k^2 \\
& - \frac{\gamma_{k-1}}{\Delta x}\int_{x_{k-1}}^{x_k} \frac{\partial R}{\partial \phi_X}\Big(\gamma_{k-1}, y_{k-1}\eta_{k-1}(x) + y_k\eta_k(x)\Big)\,dx \\
& + \frac{\gamma_k}{\Delta x}\int_{x_k}^{x_{k+1}} \frac{\partial R}{\partial \phi_X}\Big(\gamma_k, y_k\eta_k(x) + y_{k+1}\eta_{k+1}(x)\Big)\,dx \\
& + \frac{1}{\Delta x}\int_{x_{k-1}}^{x_k} R\Big(\gamma_{k-1}, y_{k-1}\eta_{k-1}(x) + y_k\eta_k(x)\Big)\,dx \\
& - \frac{1}{\Delta x}\int_{x_k}^{x_{k+1}} R\Big(\gamma_k, y_k\eta_k(x) + y_{k+1}\eta_{k+1}(x)\Big)\,dx.
\end{aligned} \tag{68}$$

We are now ready to consider the generalized Hamiltonian equation

$$i_Z\tilde{\Omega}_N = d\tilde{E}_N, \tag{69}$$

which we solve for the vector field $Z = \sum_{i=1}^{2N} \alpha_i\,\partial/\partial q_i + \beta_i\,\partial/\partial \dot{q}_i$. In the matrix representation this equation takes the form

$$\tilde{\Omega}_N^T(q,\dot{q}) \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = d\tilde{E}_N^T(q,\dot{q}). \tag{70}$$

Equations of this form are called (quasilinear) implicit ODEs (see [17], [18]). If the symplectic form is nonsingular in a neighborhood of $(q^{(0)}, \dot{q}^{(0)})$, then the equation can be solved directly via

$$Z = [\tilde{\Omega}_N^T(q, \dot{q})]^{-1} d\tilde{E}_N^T(q, \dot{q})$$

to obtain the standard explicit ODE form and standard existence/uniqueness theorems (Picard's, Peano's, etc.) of ODE theory can be invoked to show local existence and uniqueness of the flow of Z in a neighborhood of $(q^{(0)}, \dot{q}^{(0)})$. If, however, the symplectic form is singular at $(q^{(0)}, \dot{q}^{(0)})$, then there are two possibilities. The first case is

$$d\tilde{E}_N^T(q^{(0)}, \dot{q}^{(0)}) \notin \text{Range } \tilde{\Omega}_N^T(q^{(0)}, \dot{q}^{(0)}) \tag{71}$$

and it means there is no solution for Z at $(q^{(0)}, \dot{q}^{(0)})$. This type of singularity is called an *algebraic* one and it leads to so called *impasse points* (see [19]-[17], [18]).

The other case is

$$d\tilde{E}_N^T(q^{(0)}, \dot{q}^{(0)}) \in \text{Range } \tilde{\Omega}_N^T(q^{(0)}, \dot{q}^{(0)}) \tag{72}$$

and it means that there exists a nonunique solution Z at $(q^{(0)}, \dot{q}^{(0)})$. This type of singularity is called a *geometric* one. If $(q^{(0)}, \dot{q}^{(0)})$ is a limit of regular points of (70) (i.e. points where the symplectic form is nonsingular), then there might exist an integral curve of Z passing through $(q^{(0)}, \dot{q}^{(0)})$. See [19], [20], [21], [22], [23], [17], [18] for more details.

Proposition 8. *The singularities of the symplectic form $\tilde{\Omega}_N(q, \dot{q})$ are geometric.*

Proof. Suppose that the mass matrix (and thus the symplectic form) is singular at $(q^{(0)}, \dot{q}^{(0)})$. Using the block structures (60) and (65) we can write (70) as the system

$$\begin{aligned} -\tilde{\Delta}_N(q^{(0)}, \dot{q}^{(0)})\, \alpha - \tilde{M}_N(q^{(0)})\, \beta &= \xi, \\ \tilde{M}_N(q^{(0)})\, \alpha &= \tilde{M}_N(q^{(0)})\, \dot{q}^{(0)}. \end{aligned} \tag{73}$$

The second equation implies that there exists a solution $\alpha = \dot{q}^{(0)}$. In fact this is the only solution we are interested in, since it satisfies the second order condition: the Euler-Lagrange equations underlying the variationl principle are second order, so we are only interested in solutions of the form $Z = \sum_{i=1}^{2N} \dot{q}_i\, \partial/\partial q_i + \beta_i\, \partial/\partial \dot{q}_i$. The first equation can be rewritten as

$$\tilde{M}_N(q^{(0)})\, \beta = -\xi - \tilde{\Delta}_N(q^{(0)}, \dot{q}^{(0)})\, \dot{q}^{(0)}. \tag{74}$$

Since the mass matrix is singular, we must have $\gamma_{k-1} = \gamma_k$ for some k. As we saw in Section 3.3, this means that the two rows of the k^{th} 'block row' of the mass matrix (i.e., the rows containing the blocks B_{k-1}, A_k and B_k) are not linearly independent. In fact we have

$$(B_{k-1})_{2*} = -\gamma_k (B_{k-1})_{1*}, \qquad (A_k)_{2*} = -\gamma_k (A_k)_{1*}, \qquad (B_k)_{2*} = -\gamma_k (B_k)_{1*}, \tag{75}$$

where a_{m*} denotes the m^{th} row of the matrix a. Equation (74) will have a solution for β iff the RHS satisfies a similar scaling condition in the the k^{th} 'block element'. Using formulas (62), (67) and (68), we show that $-\xi - \tilde{\Delta}_N\, \dot{q}^{(0)}$ indeed has this property. Hence, $d\tilde{E}_N^T(q^{(0)}, \dot{q}^{(0)}) \in \text{Range } \tilde{\Omega}_N^T(q^{(0)}, \dot{q}^{(0)})$ and $(q^{(0)}, \dot{q}^{(0)})$ is a geometric singularity. Moreover, since $\gamma_{k-1} = \gamma_k$ defines a hypersurface in $\mathbb{R}^{2N} \times \mathbb{R}^{2N}$, $(q^{(0)}, \dot{q}^{(0)})$ is a limit of regular points.

□

Remark I.

Numerical time integration of the semi-discrete equations of motion (70) has to deal with the singularity points of the symplectic form. While there are some numerical algorithms allowing one to get past singular hypersurfaces (see [17]), it might not be very practical from the application point of view. Note that, unlike in the continuous case, the time evolution of the meshpoints X_i's is governed by the equations of motion, so the user does not have any influence on how the mesh is adapted. More importantly, there is no built-in mechanism that would prevent mesh tangling. Some preliminary numerical experiments show that the mesh points eventually collapse when started with nonzero initial velocities.

Remark II.

The singularities of the mass matrix (47) bear some similarities to the singularities of the mass matrices encountered in the Moving Finite Element method. In [24] and [25] the authors proposed introducing a small 'internodal' viscosity which penalizes the method for relative motion between the nodes and thus regularizes the mass matrix. A similar idea could be applied in our case: one could add some small ε kinetic terms to the Lagrangian (44) in order to regularize the Legendre Transform. In light of the remark made above, we did not follow this idea further and decided to take a different route instead, as described in the following sections. However, investigating further similarities between our variational approach and the Moving Finite Element method might be worthwhile. There also might be some connection to the r-adaptive method presented in [26]: the evolution of the mesh in that method is also set by the equations of motion, although the authors considered a different variational principle and different theoretical reasoning to justify the validity of their approach.

3.5. Constraints and adaptation strategy

As we saw in Section 3.4, upon discretization we lose the arbitrariness of $X(x,t)$ and the evolution of $X_i(t)$ is governed by the equations of motion, while we still want to be able to select a desired mesh adaptation strategy, like (28). This could be done by augmenting the Lagrangian (44) with Lagrange multipliers corresponding to each constraint g_i. However, it is not obvious that the dynamics of the constrained system as defined would reflect in any way the behavior of the approximated system (42). We will show that the constraints can be added via Lagrange multipliers already at the continuous level (42) and the continuous system as defined can be then discretized to arrive at (44) with the desired adaptation constraints.

3.5.1. Global constraint

As mentioned before, eventually we would like to impose the constraints

$$g_i(y_1,...,y_N,X_1,...,X_N) = 0 \qquad i = 1,...,N \tag{76}$$

on the semi-discrete system (44). Let us assume that $g : \mathbb{R}^{2N} \longrightarrow \mathbb{R}^N$, $g = (g_1,...,g_N)^T$ is C^1 and 0 is a regular value of g, so that (76) defines a submanifold. To see how these constraints can be introduced at the continuous level, let us select uniformly distributed points $x_i = i \cdot \Delta x$, $i = 0,...,N+1$, $\Delta x = X_{max}/(N+1)$ and demand that the constraints

$$g_i\Big(\varphi(x_1,t),...,\varphi(x_N,t),X(x_1,t),...,X(x_N,t)\Big) = 0, \qquad i = 1,...,N \tag{77}$$

be satisfied by $\varphi(x,t)$ and $X(x,t)$. One way of imposing these constraints is solving the system

$$
\begin{aligned}
&\delta_1\tilde{S}[\varphi,X]\cdot\delta\varphi(x,t)=0 \qquad \text{for all } \delta\varphi(x,t),\\
&g_i\Big(\varphi(x_1,t),...,\varphi(x_N,t),X(x_1,t),...,X(x_N,t)\Big)=0, \qquad i=1,...,N.
\end{aligned}
\tag{78}
$$

This system consists of one Euler-Lagrange equation that corresponds to extremizing $\tilde{S}$ with respect to φ (we saw in Section 3.1 that the other Euler-Lagrange equation is not independent) and a set of constraints enforced at some pre-selected points x_i. Note, that upon finite element discretization on a mesh coinciding with the pre-selected points this system reduces to the approach presented in Section 2: we minimize the discrete action with respect to the y_i's only and supplement the resulting equations with the constraints (76).

Another way that we want to explore consists in using Lagrange multipliers. Define the auxiliary action functional

$$
\tilde{S}_C[\varphi,X,\lambda_k]=\tilde{S}[\varphi,X]-\sum_{i=1}^{N}\int_0^{T_{max}}\lambda_i(t)\cdot g_i\Big(\varphi(x_1,t),...,\varphi(x_N,t),X(x_1,t),...,X(x_N,t)\Big)\,dt. \tag{79}
$$

We will assume that the Lagrange multipliers $\lambda_i(t)$ are at least continuous in time. According to the method of Lagrange multipliers, we seek the stationary points of $\tilde{S}_C$. This leads to the following system of equations

$$
\begin{aligned}
\delta_1\tilde{S}[\varphi,X]\cdot\delta\varphi(x,t)-\sum_{i=1}^{N}\sum_{j=1}^{N}\int_0^{T_{max}}\lambda_i(t)\,\frac{\partial g_i}{\partial y_j}\,\delta\varphi(x_j,t)\,dt=0 &\qquad \text{for all } \delta\varphi(x,t),\\
\delta_2\tilde{S}[\varphi,X]\cdot\delta X(x,t)-\sum_{i=1}^{N}\sum_{j=1}^{N}\int_0^{T_{max}}\lambda_i(t)\,\frac{\partial g_i}{\partial X_j}\,\delta X(x_j,t)\,dt=0 &\qquad \text{for all } \delta X(x,t),\\
g_i\Big(\varphi(x_1,t),...,\varphi(x_N,t),X(x_1,t),...,X(x_N,t)\Big)=0, &\qquad i=1,...,N,
\end{aligned}
\tag{80}
$$

where for clarity we suppressed writing the arguments of $\frac{\partial g_i}{\partial y_j}$ and $\frac{\partial g_i}{\partial X_j}$.

Equation (78) is more intuitive, because we directly use the arbitrariness of $X(x,t)$ and simply restrict it further by imposing constraints. It is not immediately obvious how solutions of (78) and (80) relate to each other. We would like both systems to be 'equivalent' in some sense, or at least their solution sets to overlap. Let us investigate this issue in more detail.

Suppose (φ,X) satisfy (78). Then it is quite trivial to see that $(\varphi,X,\lambda_1,...,\lambda_N)$ such that $\lambda_k\equiv 0$ satisfy (80): the second equation is implied by the first one and the other equations coincide with those of (78). At this point it should be obvious that system (80) may have more solutions for φ and X than system (78).

Proposition 9. *The only solutions $(\varphi,X,\lambda_1,...,\lambda_N)$ to (80) that satisfy (78) as well are those with $\lambda_k\equiv 0$ for all k.*

Proof. Suppose $(\varphi,X,\lambda_1,...,\lambda_N)$ satisfy both (78) and (80). System (78) implies that $\delta_1\tilde{S}\cdot\delta\varphi=0$ and $\delta_2\tilde{S}\cdot\delta X=0$. Using this in system (80) gives

$$\sum_{j=1}^{N}\int_{0}^{T_{max}} dt\,\delta\varphi(x_j,t)\,\sum_{i=1}^{N}\lambda_i(t)\,\frac{\partial g_i}{\partial y_j}=0 \qquad \text{for all } \delta\varphi(x,t),$$
$$\sum_{j=1}^{N}\int_{0}^{T_{max}} dt\,\delta X(x_j,t)\,\sum_{i=1}^{N}\lambda_i(t)\,\frac{\partial g_i}{\partial X_j}=0 \qquad \text{for all } \delta X(x,t). \tag{81}$$

In particular, this has to hold for variations $\delta\varphi$ and δX such that $\delta\varphi(x_j,t)=\delta X(x_j,t)=\nu(t)\cdot\delta_{kj}$, where $\nu(t)$ is an arbitrary continuous function of time. If we further assume that for all $x\in[0,X_{max}]$ the functions $\varphi(x,.)$ and $X(x,.)$ are continuous, both $\sum_{i=1}^{N}\lambda_i(t)\,\frac{\partial g_i}{\partial y_k}$ and $\sum_{i=1}^{N}\lambda_i(t)\,\frac{\partial g_i}{\partial X_k}$ are continuous and we get

$$Dg\Big(\varphi(x_1,t),...,\varphi(x_N,t),X(x_1,t),...,X(x_N,t)\Big)^T\cdot\lambda(t)=0 \tag{82}$$

for all t, where $\lambda=(\lambda_1,...,\lambda_N)^T$ and the $N\times 2N$ matrix $Dg=\Big[\frac{\partial g_i}{\partial y_k}\;\frac{\partial g_i}{\partial X_k}\Big]_{i,k=1,...,N}$ is the derivative of g. Since we assumed that 0 is a regular value of g and the constraint $g=0$ is satisfied by φ and X, we have that for all t the matrix Dg has full rank—that is, there exists a nonsingular $N\times N$ submatrix Ξ. Then the equation $\Xi^T\lambda(t)=0$ implies $\lambda\equiv 0$.

□

We see that considering Lagrange multipliers in (79) makes sense at the continuous level. We can now perform a finite element discretization. The auxiliary Lagrangian $\tilde{L}_C: Q\times G\times W\times Z\times\mathbb{R}^N\longrightarrow\mathbb{R}$ corresponding to (79) can be written as

$$\tilde{L}_C[\varphi,X,\varphi_t,X_t,\lambda_k]=\tilde{L}[\varphi,X,\varphi_t,X_t]-\sum_{i=1}^{N}\lambda_i\cdot g_i\Big(\varphi(x_1),...,\varphi(x_N),X(x_1),...,X(x_N)\Big), \tag{83}$$

where $\tilde{L}$ is the Lagrangian of the unconstrained theory and has been defined by (38). Let us choose a uniform mesh coinciding with the pre-selected points x_i. As in Section 3.2, we consider the restriction $\tilde{L}_{CN}=\tilde{L}_C|_{Q_N\times G_N\times W_N\times Z_N\times\mathbb{R}^N}$ and we get

$$\tilde{L}_{CN}(y_i,X_j,\dot{y}_k,\dot{X}_l,\lambda_m)=\tilde{L}_N(y_i,X_j,\dot{y}_k,\dot{X}_l)-\sum_{i=1}^{N}\lambda_i\cdot g_i(y_1,...,y_N,X_1,...,X_N). \tag{84}$$

We see that the semi-discrete Lagrangian $\tilde{L}_{CN}$ is obtained from the semi-discrete Lagrangian $\tilde{L}_N$ by adding the constraints g_i directly at the semi-discrete level, which is exactly what we set out to do at the beginning of this section. However, in the semi-discrete setting we cannot expect the Lagrange multipliers to vanish for solutions of interest. This is because there is no semi-discrete counterpart of Proposition 9. On one hand, the semi-discrete version of (78) (that is, the approach presented in Section 2) does not imply that $\delta_2\tilde{S}\cdot\delta X=0$, so the above proof will not work. On the other hand, if we supplement (78) with the equation corresponding to variations of X, then the finite element discretization will not have solutions, unless the constraint functions are integrals of motion of the system described by $\tilde{L}_N(y_i,X_j,\dot{y}_k,\dot{X}_l)$, which generally is not the case. Nonetheless, it is reasonable to expect that if the continuous system (78) has a solution, then the Lagrange multipliers of the semi-discrete system (84) should remain small.

Defining constraints by Equations (77) allowed us to use the same finite element discretization for both $\tilde{L}$ and the constraints, and to prove some correspondence between the solutions of (78) and (80). However, constraints (77) are global in the sense that they depend on the values of the fields φ and X at different points in space. Moreover, these constraints do not determine unique solutions to (78) and (80), which is a little cumbersome when discussing multisymplecticity (see Section 4).

3.5.2. Local constraint

In Section 2.4 we discussed how some adaptation constraints of interest can be derived from certain partial differential equations based on the equidistribution principle, for instance equation (27). We can view these PDEs as local constraints that only depend on pointwise values of the fields φ, X and their spatial derivatives. Let $G = G(\varphi, X, \varphi_x, X_x, \varphi_{xx}, X_{xx}, ...)$ represent such a local constraint. Then, similarly to (78), we can write our control-theoretic strategy from Section 2 as

$$\begin{aligned} &\delta_1 \tilde{S}[\varphi, X] \cdot \delta\varphi(x,t) = 0 \qquad \text{for all } \delta\varphi(x,t), \\ &G(\varphi, X, \varphi_x, X_x, \varphi_{xx}, X_{xx}, ...) = 0. \end{aligned} \tag{85}$$

Note that higher order derivatives of the fields may require the use of higher degree basis functions than the ones in (9), or of finite differences instead.

The Lagrange multiplier approach consists in defining the auxiliary Lagrangian

$$\tilde{L}_C[\varphi, X, \varphi_t, X_t, \lambda] = \tilde{L}[\varphi, X, \varphi_t, X_t] - \int_0^{X_{max}} \lambda(x) \cdot G(\varphi, X, \varphi_x, X_x, \varphi_{xx}, X_{xx}, ...)\, dx. \tag{86}$$

Suppose that the pair (φ, X) satisfies (85). Then, much like in Section 3.5.1, one can easily check that the triple $(\varphi, X, \lambda \equiv 0)$ satisfies the Euler-Lagrange equations associated with (86). However, an analog of Proposition 9 does not seem to be very interesting in this case, therefore we are not proving it here.

Introducing the constraints this way is convenient, because the Lagrangian (86) then represents a constrained multisymplectic field theory with a local constraint, which makes the analysis of multisymplecticity easier (see Section 4). The disadvantage is that discretization of (86) requires mixed methods. We will use the linear finite elements (9) to discretize $\tilde{L}[\varphi, X, \varphi_t, X_t]$, but the constraint term will be approximated via finite differences. This way we again obtain the semi-discrete Lagrangian (84), where g_i represents the discretization of G at the point $x = x_i$.

In summary, the methods presented in Section 3.5.1 and Section 3.5.2 both lead to the same semi-discrete Lagrangian, but have different theoretical advantages.

3.6. DAE formulation of the equations of motion

The Lagrangian (84) can be written as

$$\tilde{L}_{CN}(q, \dot{q}, \lambda) = \frac{1}{2}\dot{q}^T \tilde{M}_N(q)\, \dot{q} - R_N(q) - \lambda^T g(q), \tag{87}$$

where

$$R_N(q) = \sum_{k=0}^{N} \int_{x_k}^{x_{k+1}} R\Big(\gamma_k, y_k \eta_k(x) + y_{k+1}\eta_{k+1}(x)\Big) \frac{X_{k+1} - X_k}{\Delta x}\, dx. \tag{88}$$

The Euler-Lagrange equations thus take the form

$$\begin{aligned} \dot{q} &= u, \\ \tilde{M}_N(q)\, \dot{u} &= f(q,u) - Dg(q)^T \lambda, \\ g(q) &= 0, \end{aligned} \tag{89}$$

where

$$f_k(q,u) = -\frac{\partial R_N}{\partial q_k} + \sum_{i,j=1}^{2N} \Big(\frac{1}{2}\frac{\partial (\tilde{M}_N)_{ij}}{\partial q_k} - \frac{\partial (\tilde{M}_N)_{ki}}{\partial q_j}\Big) u_i u_j. \tag{90}$$

System (89) is to be solved for the unknown functions $q(t)$, $u(t)$ and $\lambda(t)$. This is a DAE system of index 3, since we are lacking a differential equation for $\lambda(t)$ and the constraint equation has to be differentiated three times in order to express $\dot{\lambda}$ as a function of q, u and λ, provided that certain regularity conditions are satisfied. Let us determine these conditions. Differentiate the constraint equation with respect to time twice to obtain the acceleration level constraint

$$Dg(q)\,\dot{u} = h(q,u), \tag{91}$$

where

$$h_k(q,u) = -\sum_{i,j=1}^{2N} \frac{\partial^2 g_k}{\partial q_i \partial q_j} u_i u_j. \tag{92}$$

We can then write (91) and the second equation of (89) together as

$$\begin{pmatrix} \tilde{M}_N(q) & Dg(q)^T \\ Dg(q) & 0 \end{pmatrix} \begin{pmatrix} \dot{u} \\ \lambda \end{pmatrix} = \begin{pmatrix} f(q,u) \\ h(q,u) \end{pmatrix}. \tag{93}$$

If we could solve this equation for $\dot{u}$ and λ in terms of q and u, then we could simply differentiate the expression for λ one more time to obtain the missing differential equation, thus showing system (89) is of index 3. System (93) is solvable if its matrix is invertible. Hence, for system (89) to be of index 3 the following condition

$$\det \begin{pmatrix} \tilde{M}_N(q) & Dg(q)^T \\ Dg(q) & 0 \end{pmatrix} \neq 0 \tag{94}$$

has to be satisfied for all q or at least in a neighborhood of the points satisfying $g(q) = 0$. Note that with suitably chosen constraints this condition allows the mass matrix to be singular.

We would like to perform time integration of this mechanical system using the symplectic (variational) Lobatto IIIA-IIIB quadratures for constrained systems (see [1], [12], [27], [28], [2], [29], [30], [31]). However, due to the singularity of the Runge-Kutta coefficient matrices (a_{ij}) and $(\bar{a}_{ij})$ for the Lobatto IIIA and IIIB schemes, the assumption (94) does not guarantee that these quadratures define a unique numerical solution: the mass matrix would need to be invertible. To circumvent this numerical obstacle we resort to a trick described in [28]. We embed our mechanical system in a higher dimensional configuration space by adding slack degrees of freedom r and $\dot{r}$ and form the augmented Lagrangian $\tilde{L}_N^A$ by modifying the kinetic term of $\tilde{L}_N$ to read

$$\tilde{L}_N^A(q,r,\dot{q},\dot{r}) = \frac{1}{2} \begin{pmatrix} \dot{q}^T & \dot{r}^T \end{pmatrix} \cdot \begin{pmatrix} \tilde{M}_N(q) & Dg(q)^T \\ Dg(q) & 0 \end{pmatrix} \cdot \begin{pmatrix} \dot{q} \\ \dot{r} \end{pmatrix} - R_N(q). \tag{95}$$

Assuming (94), the augmented system has a non-singular mass matrix. If we multiply out the terms we obtain simply

$$\tilde{L}_N^A(q,r,\dot{q},\dot{r}) = \tilde{L}_N(q,\dot{q}) + \dot{r}^T Dg(q)\,\dot{q}. \tag{96}$$

This formula in fact holds for general Lagrangians, not only for (44). In addition to $g(q) = 0$ we further impose the constraint $r = 0$. Then the augmented constrained Lagrangian takes the form

$$\tilde{L}_{CN}^A(q,r,\dot{q},\dot{r},\lambda,\mu) = \tilde{L}_N(q,\dot{q}) + \dot{r}^T Dg(q)\,\dot{q} - \lambda^T g(q) - \mu^T r. \tag{97}$$

The corresponding Euler-Lagrange equations are

$$
\begin{aligned}
\dot{q} &= u, \\
\dot{r} &= w, \\
\tilde{M}_N(q)\,\dot{u} + Dg(q)^T\,\dot{w} &= f(q,u) - Dg(q)^T\,\lambda, \\
Dg(q)\,\dot{u} &= h(q,u) - \mu, \\
g(q) &= 0, \\
r &= 0.
\end{aligned} \tag{98}
$$

It is straightforward to verify that $r(t) = 0$, $w(t) = 0$, $\mu(t) = 0$ is the exact solution and the remaining equations reduce to (89), that is, the evolution of the augmented system coincides with the evolution of the original system, by construction. The advantage is that the augmented system is now regular and we can readily apply the Lobatto IIIA-IIIB method for constrained systems to compute a numerical solution. It should be intuitively clear that this numerical solution will approximate the solution of (89) as well. What is not immediately obvious is whether a variational integrator based on (96) can be interpreted as a variational integrator based on $\tilde{L}_N$. This can be elegantly justified with the help of exact constrained discrete Lagrangians. Let $\mathcal{N} \subset Q_N \times G_N$ be the constraint submanifold defined by $g(q) = 0$. The exact constrained discrete Lagrangian $\tilde{L}_N^{C,E} : \mathcal{N} \times \mathcal{N} \longrightarrow \mathbb{R}$ is defined by

$$
\tilde{L}_N^{C,E}\big(q^{(1)}, q^{(2)}\big) = \int_0^{\Delta t} \tilde{L}_N\big(q(t), \dot{q}(t)\big)\,dt, \tag{99}
$$

where $q(t)$ is the solution to the constrained Euler-Lagrange equations (89) such that it satisfies the boundary conditions $q(0) = q^{(1)}$ and $q(\Delta t) = q^{(2)}$. Note that $\mathcal{N} \times \{0\} \subset (Q_N \times G_N) \times \mathbb{R}^N$ is the constraint submanifold defined by $g(q) = 0$ and $r = 0$. Since necessarily $r^{(1)} = r^{(2)} = 0$, we can define the exact augmented constrained discrete Lagrangian $\tilde{L}_N^{A,C,E} : \mathcal{N} \times \mathcal{N} \longrightarrow \mathbb{R}$ by

$$
\tilde{L}_N^{A,C,E}\big(q^{(1)}, q^{(2)}\big) = \int_0^{\Delta t} \tilde{L}_N^{A}\big(q(t), r(t), \dot{q}(t), \dot{r}(t)\big)\,dt, \tag{100}
$$

where $q(t)$, $r(t)$ are the solutions to the augmented constrained Euler-Lagrange equations (98) such that the boundary conditions $q(0) = q^{(1)}$, $q(\Delta t) = q^{(2)}$ and $r(0) = r(\Delta t) = 0$ are satisfied.

Proposition 10. *The exact discrete Lagrangians $\tilde{L}_N^{A,C,E}$ and $\tilde{L}_N^{C,E}$ are equal.*

Proof. Let $q(t)$ and $r(t)$ be the solutions to (98) such that the boundary conditions $q(0) = q^{(1)}$, $q(\Delta t) = q^{(2)}$ and $r(0) = r(\Delta t) = 0$ are satisfied. As argued before, we in fact have $r(t) = 0$ and $q(t)$ satisfies (89) as well. By (96) we have

$$
\tilde{L}_N^{A}\big(q(t), r(t), \dot{q}(t), \dot{r}(t)\big) = \tilde{L}_N\big(q(t), \dot{q}(t)\big)
$$

for all $t \in [0, \Delta t]$, and consequently $\tilde{L}_N^{A,C,E} = \tilde{L}_N^{C,E}$.
□

This means that any discrete Lagrangian $\tilde{L}_d : (Q_N \times G_N) \times \mathbb{R}^N \times (Q_N \times G_N) \times \mathbb{R}^N \longrightarrow \mathbb{R}$ that approximates $\tilde{L}_N^{A,C,E}$ to order s also approximates $\tilde{L}_N^{C,E}$ to the same order, that is, a variational integrator for (98), in particular our Lobatto IIIA-IIIB scheme, is also a variational integrator for (89).

Backward error analysis.

The advantage of the Lagrange multiplier approach is the fact that upon spatial discretization we deal with a constrained mechanical system. Backward error analysis of symplectic/variational numerical schemes for such systems shows that the modified equations also describe a constrained mechanical system for a nearby Hamiltonian (see Theorem 5.6 in Section IX.5.2 of [1]). Therefore, we expect the Lagrange multiplier strategy to demonstrate better performance in terms of energy conservation than the control-theoretic strategy. The Lagrange multiplier approach makes better use of the geometry underlying the field theory we consider, the key idea being to *treat the reparametrization field $X(x,t)$ as an additional dynamical degree of freedom on equal footing with $\varphi(x,t)$.*

4. Multisymplectic field theory formalism

In Section 2 and Section 3 we took the view of infinite dimensional manifolds of fields as configuration spaces and presented a way to construct space-adaptive variational integrators in that formalism. We essentially applied symplectic integrators to semi-discretized Lagrangian field theories. In this section we show how r-adaptive integrators can be described in the more general framework of multisymplectic geometry. In particular we show that some of the integrators obtained in the previous sections can be interpreted as multisymplectic variational integrators. Multisymplectic geometry provides a covariant formalism for the study of field theories in which time and space are treated on equal footing, as a conseqence of which multisymplectic variational integrators allow for more general discretizations of spacetime, such that, for instance, each element of space may be integrated with a different timestep (see [4]). For the convenience of the reader, below we briefly review some background material and provide relevant references for further details. We then proceed to reformulate our adaptation strategies in the language of multisymplectic field theory.

4.1. Background material

Lagrangian mechanics and Veselov-type discretizations

Let Q be the configuration manifold of a certain mechanical system and TQ its tangent bundle. Denote the coordinates on Q by q^i, and on TQ by $(q^i, \dot{q}^i)$, where $i = 1, 2, ..., n$. The system is described by defining the Lagrangian $L : TQ \longrightarrow \mathbb{R}$ and the corresponding action functional $S[q(t)] = \int_a^b L\big(q^i(t), \dot{q}^i(t)\big)\, dt$. The dynamics is obtained through Hamilton's principle, which seeks the curves $q(t)$ for which the functional $S[q(t)]$ is stationary under variations of $q(t)$ with fixed endpoints, i.e. we seek $q(t)$ such that

$$dS[q(t)] \cdot \delta q(t) = \frac{d}{d\epsilon}\bigg|_{\epsilon=0} S[q_\epsilon(t)] = 0 \tag{101}$$

for all $\delta q(t)$ with $\delta q(a) = \delta q(b) = 0$, where $q_\epsilon(t)$ is a smooth family of curves satisfying $q_0 = q$ and $\frac{d}{d\epsilon}\big|_{\epsilon=0} q_\epsilon = \delta q$. By using integration by parts, the Euler-Lagrange equations follow as

$$\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} = 0. \tag{102}$$

The canonical symplectic form Ω on T^*Q, the $2n$-dimensional cotangent bundle of Q, is given by $\Omega = dq^i \wedge dp_i$, where summation over i is implied and (q^i, p_i) are the canonical coordinates on T^*Q. The Lagrangian defines the Legendre transformation $\mathbb{F}L : TQ \longrightarrow T^*Q$, which in coordinates is given by $(q^i, p_i) = (q^i, \frac{\partial L}{\partial \dot{q}^i})$. We then define the Lagrange 2-form on TQ by pulling back the canonical symplectic form, i.e. $\Omega_L = \mathbb{F}L^*\Omega$. If the Legendre transformation is a local diffeomorphism, then Ω_L is a symplectic form. The Lagrange vector field is a vector field X_E on TQ that satisfies $X_E \lrcorner \Omega_L = dE$, where the energy E is defined by $E(v_q) = \mathbb{F}L(v_q) \cdot v_q - L(v_q)$ and $\lrcorner$ denotes the interior product, i.e. the contraction of a differential form with a vector field. It can be shown that the flow F_t of this vector field preserves the symplectic form, that is, $F_t^*\Omega_L = \Omega_L$. The flow F_t is obtained by solving the Euler-Lagrange equations (102).

For a Veselov-type discretization we essentially replace TQ with $Q \times Q$, which serves as a discrete approximation of the tangent bundle. We define a discrete Lagrangian L_d as a smooth map $L_d : Q \times Q \longrightarrow \mathbb{R}$ and the corresponding discrete action $S = \sum_{k=0}^{N-1} L_d(q_k, q_{k+1})$. The variational principle now seeks a sequence $q_0, q_1, ..., q_N$ that extremizes S for variations holding the endpoints q_0 and q_N fixed. The Discrete Euler-Lagrange equations follow

$$D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}) = 0. \tag{103}$$

This implicitly defines a discrete flow $F : Q \times Q \longrightarrow Q \times Q$ such that $F(q_{k-1}, q_k) = (q_k, q_{k+1})$. One can define the discrete Lagrange 2-form on $Q \times Q$ by $\omega_L = \frac{\partial^2 L_d}{\partial q_0^i \partial q_1^j} dq_0^i \wedge dq_1^j$, where (q_0^i, q_1^j) denotes the coordinates on $Q \times Q$. It then follows that the discrete flow F is symplectic, i.e. $F^*\omega_L = \omega_L$.

Given a continuous Lagrangian system with $L : TQ \longrightarrow \mathbb{R}$ one chooses a corresponding discrete Lagrangian as an approximation $L_d(q_k, q_{k+1}) \approx \int_{t_k}^{t_{k+1}} L\big(q(t), \dot{q}(t)\big)\, dt$, where $q(t)$ is the solution of the Euler-Lagrange equations corresponding to L with the boundary values $q(t_k) = q_k$ and $q(t_{k+1}) = q_{k+1}$.

For more details regarding Lagrangian mechanics, variational principles, and symplectic geometry, see [32]. Discrete Mechanics and variational integrators are discussed in [2].

Multisymplectic geometry and Lagrangian field theory

Let $\mathcal{X}$ be an oriented manifold representing the $(n+1)$-dimensional spacetime with local coordinates $(x^0, x^1, \ldots, x^n) \equiv (t, x)$, where $x^0 \equiv t$ is time and $(x^1, \ldots, x^n) \equiv x$ are space coordinates. Physical fields are sections of a configuration fiber bundle $\pi_{\mathcal{X}Y} : Y \longrightarrow \mathcal{X}$, that is, continuous maps $\phi : \mathcal{X} \longrightarrow Y$ such that $\pi_{\mathcal{X}Y} \circ \phi = \mathrm{id}_{\mathcal{X}}$. This means that for every $(t, x) \in \mathcal{X}$, $\phi(t, x)$ is in the fiber over (t, x), which is $Y_{(t,x)} = \pi_{\mathcal{X}Y}^{-1}((t, x))$. The evolution of the field takes place on the first jet bundle J^1Y, which is the analog of TQ for mechanical systems. J^1Y is defined as the affine bundle over Y such that for $y \in Y_{(t,x)}$ the fiber $J^1_y Y$ consists of linear maps $\vartheta : T_{(t,x)}\mathcal{X} \rightarrow T_y Y$ satisfying the condition $T\pi_{\mathcal{X}Y} \circ \vartheta = \mathrm{id}_{T_{(t,x)}\mathcal{X}}$. The local coordinates (x^μ, y^a) on Y induce the coordinates $(x^\mu, y^a, v^a{}_\mu)$ on J^1Y. Intuitively, the first jet bundle consists of the configuration bundle Y, and of the first partial derivatives of the field variables with respect to the independent variables. Let $\phi(x^0, \ldots, x^n) = (x^0, \ldots, x^n, y^1, \ldots, y^m)$ in coordinates and let $v^a{}_\mu = y^a{}_{,\mu} = \partial y^a / \partial x^\mu$ denote the partial derivatives. We can think of J^1Y as a fiber bundle over $\mathcal{X}$. Given a section $\phi : \mathcal{X} \longrightarrow Y$, we can define its first jet prolongation $j^1\phi : \mathcal{X} \longrightarrow J^1Y$, in coordinates given by $j^1\phi(x^0, x^1, \ldots, x^n) = (x^0, x^1, \ldots, x^n, y^1, \ldots, y^m, y^1{}_{,0}, \ldots, y^m{}_{,n})$, which is a section of the fiber bundle J^1Y over $\mathcal{X}$. For higher order field theories we consider higher order jet bundles, defined iteratively by $J^2Y = J^1(J^1Y)$ and so on. The local coordinates on J^2Y are denoted $(x^\mu, y^a, v^a{}_\mu, w^a{}_\mu, \kappa^a_{\mu\nu})$. The second jet prolongation $j^2\phi : \mathcal{X} \longrightarrow J^2Y$ is given in coordinates by $j^2\phi(x^\mu) = (x^\mu, y^a, y^a{}_{,\mu}, y^a{}_{,\mu}, y^a{}_{,\mu,\nu})$.

Lagrangian density for first order field theories is defined as a map $\mathcal{L} : J^1Y \longrightarrow \mathbb{R}$. The corresponding action functional is $S[\phi] = \int_{\mathcal{U}} \mathcal{L}(j^1\phi)\, d^{n+1}x$, where $\mathcal{U} \subset \mathcal{X}$. Hamilton's principle seeks fields $\phi(t, x)$ that extremize S, that is

$$\frac{d}{d\lambda}\bigg|_{\lambda=0} S[\eta_Y^\lambda \circ \phi] = 0 \tag{104}$$

for all η_Y^λ that keep the boundary conditions on $\partial\mathcal{U}$ fixed, where $\eta_Y^\lambda : Y \longrightarrow Y$ is the flow of a vertical vector field V on Y. This leads to the Euler-Lagrange equations

$$\frac{\partial \mathcal{L}}{\partial y^a}(j^1\phi) - \frac{\partial}{\partial x^\mu}\left(\frac{\partial \mathcal{L}}{\partial v^a{}_\mu}(j^1\phi)\right) = 0. \tag{105}$$

Given the Lagrangian density $\mathcal{L}$ one can define the Cartan $(n+1)$-form $\Theta_{\mathcal{L}}$ on J^1Y, in local coordinates given by $\Theta_{\mathcal{L}} = \frac{\partial \mathcal{L}}{\partial v^a{}_\mu} dy^a \wedge d^n x_\mu + (\mathcal{L} - \frac{\partial \mathcal{L}}{\partial v^a{}_\mu} v^a{}_\mu) d^{n+1}x$, where $d^n x_\mu = \partial_\mu \lrcorner\, d^{n+1}x$. The multisymplectic $(n+2)$-form is then defined by $\Omega_{\mathcal{L}} = -d\Theta_{\mathcal{L}}$. Let $\mathcal{P}$ be the set of solutions of the Euler-Lagrange equations, that is, the set of sections ϕ satisfying (104) or (105). For a given $\phi \in \mathcal{P}$, let $\mathcal{F}$ be the set of first variations, that is, the set of vector fields V on J^1Y such that $(t,x) \to \eta_Y^\epsilon \circ \phi(t,x)$ is also a solution, where η_Y^ϵ is the flow of V. The multisymplectic form formula states that if $\phi \in \mathcal{P}$ then for all V and W in $\mathcal{F}$,

$$\int_{\partial\mathcal{U}} (j^1\phi)^* (j^1 V \lrcorner j^1 W \lrcorner \Omega_{\mathcal{L}}) = 0, \tag{106}$$

where j^1V is the jet prolongation of V, that is, the vector field on J^1Y in local coordinates given by $j^1V = (V^\mu, V^a, \frac{\partial V^a}{\partial x^\mu} + \frac{\partial V^a}{\partial y^b} v^b{}_\mu - v^a{}_\nu \frac{\partial V^\nu}{\partial x^\mu})$, where $V = (V^\mu, V^a)$ in local coordinates. The multisymplectic form formula is the multisymplectic counterpart of the fact that in finite-dimensional mechanics, the flow of a mechanical system consists of symplectic maps.

For a k^{th}-order Lagrangian field theory with the Lagrangian density $\mathcal{L} : J^kY \longrightarrow \mathbb{R}$, analogous geometric structures are defined on $J^{2k-1}Y$. In particular, for a second-order field theory the multisymplectic $(n+2)$-form $\Omega_{\mathcal{L}}$ is defined on J^3Y and a similar multisymplectic form formula can be proven. If the Lagrangian density does not depend on the second order time derivatives of the field, it is convenient to define the subbundle $J_0^2Y \subset J^2Y$ such that $J_0^2Y = \{\vartheta \in J^2Y \,|\, \kappa^a_{00} = 0\}$.

For more information about the geometry of jet bundles, see [33]. The multisymplectic formalism in field theory is discussed in [34]. The multisymplectic form formula for first-order field theories is derived in [3], and generalized for second-order field theories in [35]. Higher order field theory is considered in [36].

Multisymplectic variational integrators

Veselov-type discretization can be generalized to multisymplectic field theory. We take $\mathcal{X} = \mathbb{Z} \times \mathbb{Z} = \{(j,i)\}$, where for simplicity we consider $\dim \mathcal{X} = 2$, i.e. $n = 1$. The configuration fiber bundle is $Y = \mathcal{X} \times \mathscr{F}$ for some smooth manifold $\mathscr{F}$. The fiber over $(j,i) \in \mathcal{X}$ is denoted Y_{ji} and its elements y_{ji}. A rectangle $\square$ of $\mathcal{X}$ is an ordered 4-tuple of the form $\square = \big((j,i),(j,i+1),(j+1,i+1),(j+1,i)\big) = (\square^1,\square^2,\square^3,\square^4)$. The set of all rectangles in $\mathcal{X}$ is denoted $\mathcal{X}^\square$. A point (j,i) is touched by a rectangle if it is a vertex of that rectangle. Let $\mathcal{U} \subset \mathcal{X}$. Then $(j,i) \in \mathcal{U}$ is an interior point of $\mathcal{U}$ if $\mathcal{U}$ contains all four rectangles that touch (j,i). The interior $\mathrm{int}\,\mathcal{U}$ is the set of all interior points of $\mathcal{U}$. The closure $\mathrm{cl}\,\mathcal{U}$ is the union of all rectangles touching interior points of $\mathcal{U}$. The boundary of $\mathcal{U}$ is defined by $\partial\mathcal{U} = (\mathcal{U} \cap \mathrm{cl}\,\mathcal{U}) \backslash \mathrm{int}\,\mathcal{U}$. A section of Y is a map $\phi : \mathcal{U} \subset \mathcal{X} \to Y$ such that $\phi(j,i) \in Y_{ji}$. We can now define the discrete first jet bundle of Y as

$$\begin{aligned} J^1Y &= \big\{(y_{ji}, y_{j\,i+1}, y_{j+1\,i+1}, y_{j+1\,i}) \;\big|\; (j,i) \in \mathcal{X},\, y_{ji}, y_{j\,i+1}, y_{j+1\,i+1}, y_{j+1\,i} \in \mathscr{F}\big\} \\ &= \mathcal{X}^\square \times \mathscr{F}^4. \end{aligned} \tag{107}$$

Intuitively, the discrete first jet bundle is the set of all rectangles together with four values assigned to their vertices. Those four values are enough to approximate the first derivatives of a smooth section with respect to time and space using, for instance, finite differences. The first jet prolongation of a section ϕ of Y is the map $j^1\phi : \mathcal{X}^\square \to J^1Y$ defined by $j^1\phi(\square) = (\square, \phi(\square^1), \phi(\square^2), \phi(\square^3), \phi(\square^4))$. For a vector field V on Y, let V_{ji} be its restriction to Y_{ji}. Define a discrete Lagrangian $L : J^1Y \to \mathbb{R}$, $L = L(y_1, y_2, y_3, y_4)$, where for convenience we omit writing the base rectangle. The associated discrete action is given by

$$S[\phi] = \sum_{\square \subset \mathcal{U}} L \circ j^1\phi(\square).$$

The discrete variational principle seeks sections that extremize the discrete action, that is, mappings $\phi(j,i)$ such that

$$\frac{d}{d\lambda}\bigg|_{\lambda=0} S[\phi_\lambda] = 0 \tag{108}$$

for all vector fields V on Y that keep the boundary conditions on $\partial\mathcal{U}$ fixed, where $\phi_\lambda(j,i) = F_\lambda^{V_{ji}}(\phi(j,i))$ and $F_\lambda^{V_{ji}}$ is the flow of V_{ji} on $\mathscr{F}$. This is equivalent to the discrete Euler-Lagrange equations

$$\begin{aligned}\frac{\partial L}{\partial y_1}(y_{ji}, y_{j\,i+1}, y_{j+1\,i+1}, y_{j+1\,i}) + \frac{\partial L}{\partial y_2}(y_{j\,i-1}, y_{ji}, y_{j+1\,i}, y_{j+1\,i-1}) +\\ + \frac{\partial L}{\partial y_3}(y_{j-1\,i-1}, y_{j-1\,i}, y_{ji}, y_{j\,i-1}) + \frac{\partial L}{\partial y_4}(y_{j-1\,i}, y_{j-1\,i+1}, y_{j\,i+1}, y_{ji}) = 0\end{aligned} \tag{109}$$

for all $(j,i) \in \operatorname{int}\mathcal{U}$, where we adopt the convention $\phi(j,i) = y_{ji}$. In analogy to the Veselov discretization of mechanics, we can define four 2-forms Ω_L^l on J^1Y, where $l = 1,2,3,4$ and $\Omega_L^1 + \Omega_L^2 + \Omega_L^3 + \Omega_L^4 = 0$, that is, only three 2-forms of these forms are independent. The 4-tuple $(\Omega_L^1, \Omega_L^2, \Omega_L^3, \Omega_L^4)$ is the discrete analog of the multisymplectic form $\Omega_\mathcal{L}$. We refer the reader to the literature for details, e.g. [3]. By analogy to the continuous case, let $\mathcal{P}$ be the set of solutions of the discrete Euler-Lagrange equations (109). For a given $\phi \in \mathcal{P}$, let $\mathcal{F}$ be the set of first variations, that is, the set of vector fields V on J^1Y defined similarly as in the continuous case. The discrete multisymplectic form formula then states that if $\phi \in \mathcal{P}$ then for all V and W in $\mathcal{F}$,

$$\sum_{\substack{\square \\ \square\cap\mathcal{U}\neq\emptyset}} \left(\sum_{\substack{l \\ \square^l\in\partial\mathcal{U}}} \Big[(j^1\phi)^*(j^1V \lrcorner\, j^1W \lrcorner\, \Omega_L^l)\Big](\square) \right) = 0, \tag{110}$$

where the jet prolongations are defined to be

$$j^1V(y_{\square^1}, y_{\square^2}, y_{\square^3}, y_{\square^4}) = \big(V_{\square^1}(y_{\square^1}), V_{\square^2}(y_{\square^2}), V_{\square^3}(y_{\square^3}), V_{\square^4}(y_{\square^4})\big). \tag{111}$$

The discrete form formula (110) is in direct analogy to the multisymplectic form formula (106) that holds in the continuous case.

Given a continuous Lagrangian density $\mathcal{L}$ one chooses a corresponding discrete Lagrangian as an approximation $L(y_{\square^1}, y_{\square^2}, y_{\square^3}, y_{\square^4}) \approx \int_{\overline{\square}} \mathcal{L} \circ j^1\bar{\phi}\, dx\, dt$, where $\overline{\square}$ is the rectangular region of the continuous spacetime that contains $\square$ and $\bar{\phi}(t,x)$ is the solution of the Euler-Lagrange equations corresponding to $\mathcal{L}$ with the boundary values at the vertices of $\square$ corresponding to $y_{\square^1}$, $y_{\square^2}$, $y_{\square^3}$, and $y_{\square^4}$.

The discrete second jet bundle J^2Y can be defined by considering ordered 9-tuples

$$\begin{aligned}\boxplus &= \big((j-1,i-1),(j-1,i),(j-1,i+1),(j,i-1),\\ &\qquad (j,i),(j,i+1),(j+1,i-1),(j+1,i),(j+1,i+1)\big)\\ &= (\boxplus^1, \boxplus^2, \boxplus^3, \boxplus^4, \boxplus^5, \boxplus^6, \boxplus^7, \boxplus^8, \boxplus^9)\end{aligned} \tag{112}$$

instead of rectangles $\square$, and the discrete subbundle J_0^2Y can be defined by considering 6-tuples

$$\begin{aligned}\square\!\square &= \big((j,i-1),(j,i),(j,i+1),(j+1,i+1),(j+1,i),(j+1,i-1)\big)\\ &= (\square\!\square^1, \square\!\square^2, \square\!\square^3, \square\!\square^4, \square\!\square^5, \square\!\square^6).\end{aligned} \tag{113}$$

Similar constructions then follow and a similar discrete multisymplectic form formula can be derived for a second order field theory.

Multisymplectic variational integrators for first order field theories are introduced in [3], and generalized for second-order field theories in [35].

4.2. Analysis of the control-theoretic approach

Continuous setting

We now discuss a multisymplectic setting for the approach presented in Section 2. Let the computational spacetime be $\mathcal{X} = \mathbb{R} \times \mathbb{R}$ with coordinates (t, x) and consider the trivial configuration bundle $Y = \mathcal{X} \times \mathbb{R}$ with coordinates (t, x, y). Let $\mathcal{U} = [0, T_{max}] \times [0, X_{max}]$ and let our scalar field be represented by a section $\tilde{\varphi} : \mathcal{U} \longrightarrow Y$ with the coordinate representation $\tilde{\varphi}(t, x) = (t, x, \varphi(t, x))$. Let (t, x, y, v_t, v_x) denote local coordinates on $J^1 Y$. In these coordinates the first jet prolongation of $\tilde{\varphi}$ is represented by $j^1\tilde{\varphi}(t, x) = (t, x, \varphi(t, x), \varphi_t(t, x), \varphi_x(t, x))$. Then the Lagrangian density (6) can be viewed as a mapping $\tilde{\mathcal{L}} : J^1 Y \longrightarrow \mathbb{R}$. The corresponding action (3) can now be expressed as

$$\tilde{S}[\tilde{\varphi}] = \int_{\mathcal{U}} \tilde{\mathcal{L}}(j^1\tilde{\varphi})\, dt \wedge dx, \tag{114}$$

Just like in Section 2, let us for the moment assume that the function $X : \mathcal{U} \longrightarrow [0, X_{max}]$ is known, so that we can view $\tilde{\mathcal{L}}$ as being time and space dependent. The dynamics is obtained by extremizing $\tilde{S}$ with respect to $\tilde{\varphi}$, that is, by solving for $\tilde{\varphi}$ such that

$$\frac{d}{d\lambda}\bigg|_{\lambda=0} \tilde{S}[\eta_Y^\lambda \circ \tilde{\varphi}] = 0 \tag{115}$$

for all η_Y^λ that keep the boundary conditions on $\partial\mathcal{U}$ fixed, where $\eta_Y^\lambda : Y \longrightarrow Y$ is the flow of a vertical vector field V on Y. Therefore, for an *a priori* known $X(t, x)$ the multisymplectic form formula (106) is satisfied for solutions of (115).

Consider the additional bundle $\pi_{\mathcal{X}\mathcal{B}} : \mathcal{B} = \mathcal{X} \times [0, X_{max}] \longrightarrow \mathcal{X}$ whose sections $\tilde{X} : \mathcal{U} \longrightarrow \mathcal{B}$ represent our diffeomorphisms. Let $\tilde{X}(t, x) = (t, x, X(t, x))$ denote a local coordinate representation and assume $X(t, .)$ is a diffeomorphism. Then define $\tilde{Y} = Y \oplus \mathcal{B}$. We have $J^k\tilde{Y} \cong J^k Y \oplus J^k\mathcal{B}$. In Section 3.5.2 we argued that the moving mesh partial differential equation (25) can be interpreted as a local constraint on the fields $\tilde{\varphi}$, $\tilde{X}$ and their spatial derivatives. This constraint can be represented by a function $G : J^k\tilde{Y} \longrightarrow \mathbb{R}$. Sections $\tilde{\varphi}$ and $\tilde{X}$ satisfy the constraint if $G(j^k\tilde{\varphi}, j^k\tilde{X}) = 0$. Therefore our control-theoretic strategy expressed in equations (85) can be rewritten as

$$\begin{aligned} \frac{d}{d\lambda}\bigg|_{\lambda=0} \tilde{S}[\eta_Y^\lambda \circ \tilde{\varphi}] &= 0, \\ G(j^k\tilde{\varphi}, j^k\tilde{X}) &= 0, \end{aligned} \tag{116}$$

for all η_Y^λ, similarly as above. Let us argue how to interpret the notion of multisymplecticity for this problem. Intuitively, multisymplecticity should be understood in a sense similar to Proposition 3. We first solve the problem (116) for $\tilde{\varphi}$ and $\tilde{X}$, given some initial and boundary conditions. Then we substitute this $\tilde{X}$ into the problem (115). Let $\mathcal{P}$ be the set of solutions to this problem. Naturally, $\tilde{\varphi} \in \mathcal{P}$. The multisymplectic form formula (106) will be satisfied for all fields in $\mathcal{P}$, but the constraint $G = 0$ will be satisfied only for $\tilde{\varphi}$.

Discretization

Discretize the computational spacetime $\mathbb{R} \times \mathbb{R}$ by picking the discrete set of points $t_j = j \cdot \Delta t$, $x_i = i \cdot \Delta x$, and define $\mathcal{X} = \{(j, i)\,|\, j, i \in \mathbb{Z}\}$. Let $\mathcal{X}^{\square}$ and $\mathcal{X}^{\boxplus}$ be the set of rectangles and 6-tuples in $\mathcal{X}$, respectively. The discrete configuration bundle is $Y = \mathcal{X} \times \mathbb{R}$ and for convenience of notation let the elements of the fiber Y_{ji} be denoted by y_i^j. Let $\mathcal{U} = \{(j, i)\,|\, j = 0, 1, \ldots, M+1,\ i = 0, 1, \ldots, N+1\}$, where $\Delta x = X_{max}/(N+1)$ and $\Delta t = T_{max}/(M+1)$. Suppose we have a discrete Lagrangian $\tilde{L} : J^1 Y \longrightarrow \mathbb{R}$ and the corresponding discrete action $\tilde{S}$ that approximates (114), where we assume that $X(t,x)$ is known and of the form (10). A variational integrator is obtained by solving

$$\frac{d}{d\lambda}\bigg|_{\lambda=0} \tilde{S}[\tilde{\varphi}_\lambda] = 0 \tag{117}$$

for a discrete section $\tilde{\varphi} : \mathcal{U} \longrightarrow Y$, as described in Section 4.1. This integrator is multisymplectic, i.e. the discrete multisymplectic form formula (110) is satisfied.

Example: Midpoint rule.

In (20) consider the 1-stage symplectic partitioned Runge-Kutta method with the coefficients $a_{11} = \bar{a}_{11} = c_1 = 1/2$ and $b_1 = \bar{b}_1 = 1$. This method is often called the *midpoint rule* and is a 2-nd order member of the Gauss family of quadratures. It can be easily shown (see [1], [2]) that the discrete Lagrangian (15) for this method is given by

$$\tilde{L}_d(t_j, y^j, t_{j+1}, y^{j+1}) = \Delta t \cdot \tilde{L}_N\bigg(\frac{y^j + y^{j+1}}{2}, \frac{y^{j+1} - y^j}{\Delta t}, t_j + \frac{1}{2}\Delta t\bigg), \tag{118}$$

where $\Delta t = t_{j+1} - t_j$ and $y^j = (y^j_1, \ldots, y^j_N)$. Using (5) and (13) we can write

$$\tilde{L}_d(t_j, y^j, t_{j+1}, y^{j+1}) = \sum_{i=0}^{N} \tilde{L}(y^j_i, y^j_{i+1}, y^{j+1}_{i+1}, y^{j+1}_i), \tag{119}$$

where we defined the discrete Lagrangian $\tilde{L} : J^1 Y \longrightarrow \mathbb{R}$ by the formula

$$\tilde{L}(y^j_i, y^j_{i+1}, y^{j+1}_{i+1}, y^{j+1}_i) = \Delta t \int_{x_i}^{x_{i+1}} \tilde{\mathcal{L}}\Big(\bar{\varphi}(x), \bar{\varphi}_x(x), \bar{\varphi}_t(x), x, t_j + \frac{1}{2}\Delta t\Big)\, dx \tag{120}$$

with

$$\begin{aligned}
\bar{\varphi}(x) &= \frac{y^j_i + y^{j+1}_i}{2}\eta_i(x) + \frac{y^j_{i+1} + y^{j+1}_{i+1}}{2}\eta_{i+1}(x), \\
\bar{\varphi}_x(x) &= \frac{1}{2}\frac{y^j_{i+1} - y^j_i}{\Delta x} + \frac{1}{2}\frac{y^{j+1}_{i+1} - y^{j+1}_i}{\Delta x}, \\
\bar{\varphi}_t(x) &= \frac{y^{j+1}_i - y^j_i}{\Delta t}\eta_i(x) + \frac{y^{j+1}_{i+1} - y^j_{i+1}}{\Delta t}\eta_{i+1}(x).
\end{aligned} \tag{121}$$

Given the Lagrangian density $\tilde{\mathcal{L}}$ as in (6), and assuming $X(t,x)$ is known, one can evaluate the integral in (120) explicitly. It is now a straightforward calculation to show that the discrete variational principle (117) for the discrete Lagrangian $\tilde{L}$ as defined is equivalent to the Discrete Euler-Lagrange equations (103) for $\tilde{L}_d$, and consequently to (20).

This shows that the 2-nd order Gauss method applied to (20) defines a multisymplectic method in the sense of formula (110). However, for other symplectic partitioned Runge-Kutta methods of interest to us, namely the 4-th order Gauss and the 2-nd/4-th order Lobatto IIIA-IIIB methods, it is not possible to isolate a discrete Lagrangian $\tilde{L}$ that would only depend on four values y^j_i, y^j_{i+1}, y^{j+1}_{i+1}, y^{j+1}_i. The mentioned methods have more internal stages, and the equations (20) couple them in a nontrivial way. Effectively, at any given time step the internal stages depend on all the values $y^j_1, \ldots, y^j_N$ and y^{j+1}_1, $\ldots, y^{j+1}_N$, and it it not possible to express the discrete Lagrangian (15) as a sum similar to (119). The resulting integrators are still variational, since they are derived by applying the discrete variational principle (117) to some discrete action $\tilde{S}$, but this action cannot be expressed as the sum of $\tilde{L}$ over all rectangles. Therefore, these integrators are not multisymplectic, at least not in the sense of formula (110).

Constraints.

Let the additional bundle be $\mathcal{B} = \mathcal{X} \times [0, X_{max}]$ and denote by X^n_j the elements of the fiber $\mathcal{B}_{ji}$. Define $\tilde{Y} = Y \oplus \mathcal{B}$. We have $J^k\tilde{Y} \cong J^kY \oplus J^k\mathcal{B}$. Suppose $G : J^k\tilde{Y} \longrightarrow \mathbb{R}$ represents a discretization of the continuous constraint. For instance, one can enforce a uniform mesh by defining $G : J^1\tilde{Y} \to \mathbb{R}$, $G(j^1\tilde{\varphi}, j^1\tilde{X}) = X_x - 1$ at the continuous level. The discrete counterpart will be defined on the discrete jet bundle $J^1\tilde{Y}$ by the formula

$$G(y^j_i, y^j_{i+1}, y^{j+1}_{i+1}, y^{j+1}_i, X^j_i, X^j_{i+1}, X^{j+1}_{i+1}, X^{j+1}_i) = \frac{X^j_{i+1} - X^j_i}{\Delta x} - 1. \tag{122}$$

Arc-length equidistribution can be realized by enforcing (27), that is, $G : J^2_0\tilde{Y} \to \mathbb{R}$, $G(j^2_0\tilde{\varphi}, j^2_0\tilde{X}) = \alpha^2\varphi_x\varphi_{xx} + X_xX_{xx}$. The discrete counterpart will be defined on the discrete subbundle $J^2_0\tilde{Y}$ by the formula

$$G(y_{\boxbar^l}, X_{\boxbar^r}) = \alpha^2(y_{\boxbar^3} - y_{\boxbar^2})^2 + (X_{\boxbar^3} - X_{\boxbar^2})^2 - \alpha^2(y_{\boxbar^2} - y_{\boxbar^1})^2 - (X_{\boxbar^2} - X_{\boxbar^1})^2, \tag{123}$$

where for convenience we used the notation introduced in (113) and $l, r = 1, \ldots, 6$. Note that (123) coincides with (28). In fact, g_i in (28) is nothing else but G computed on an element of $J^2_0\tilde{Y}$ over the base 6-tuple $\boxbar$ such that $\boxbar^2 = (j, i)$. The only difference is that in (28) we assumed g_i might depend on *all* the field values at a given time step, while G only takes arguments *locally*, i.e. it depends on *at most* 6 field values on a given 6-tuple.

A numerical scheme is now obtained by simultaneously solving the discrete Euler-Lagrange equations (109) resulting from (117) and the equation $G = 0$. If we know y^{j-1}_i, X^{j-1}_i, y^j_i and X^j_i for $i = 1, \ldots, N$, this system of equations allows us to solve for y^{j+1}_i, X^{j+1}_i. This numerical scheme is multisymplectic in the sense similar to Proposition 4. If we take $X(t, x)$ to be a sufficiently smooth interpolation of the values X^j_i and substitute it in the problem (117), then the resulting multisymplectic integrator will yield the same numerical values y^{j+1}_i.

4.3. Analysis of the Lagrange multiplier approach

Continuous setting

We now turn to describing the Lagrange multiplier approach in a multisymplectic setting. Similarly as in Section 4.2, let the computational spacetime be $\mathcal{X} = \mathbb{R} \times [0, X_{max}]$ with coordinates (t, x) and consider the trivial configuration bundles $\pi_{\mathcal{X}Y} : Y = \mathcal{X} \times \mathbb{R} \longrightarrow \mathcal{X}$ and $\pi_{\mathcal{X}\mathcal{B}} : \mathcal{B} = \mathcal{X} \times [0, X_{max}] \longrightarrow \mathcal{X}$. Let our scalar field be represented by a section $\tilde{\varphi} : \mathcal{X} \longrightarrow Y$ with the coordinate representation $\tilde{\varphi}(t, x) = (t, x, \varphi(t, x))$ and our diffeomorphism by a section $\tilde{X} : \mathcal{X} \longrightarrow \mathcal{B}$ with the local representation $\tilde{X}(t, x) = (t, x, X(t, x))$. Let the total configuration bundle be $\tilde{Y} = Y \oplus \mathcal{B}$. Then the Lagrangian density (6) can be viewed as a mapping $\tilde{\mathcal{L}} : J^1\tilde{Y} \cong J^1Y \oplus J^1\mathcal{B} \longrightarrow \mathbb{R}$. The corresponding action (3) can now be expressed as

$$\tilde{S}[\tilde{\varphi}, \tilde{X}] = \int_{\mathcal{U}} \tilde{\mathcal{L}}(j^1\tilde{\varphi}, j^1\tilde{X})\, dt \wedge dx, \tag{124}$$

where $\mathcal{U} = [0, T_{max}] \times [0, X_{max}]$. As before, the MMPDE constraint can be represented by a function $G : J^k\tilde{Y} \longrightarrow \mathbb{R}$. Two sections $\tilde{\varphi}$ and $\tilde{X}$ satisfy the constraint if

$$G(j^k\tilde{\varphi}, j^k\tilde{X}) = 0. \tag{125}$$

Vakonomic formulation.

We now face the problem of finding the right equations of motion. We want to extremize the action functional (124) in some sense, subject to the constraint (125). Note that the constraint is essentially *nonholonomic*, as it depends on the derivatives of the fields. Assuming G is a submersion, $G = 0$ defines a submanifold of $J^k\tilde{Y}$, but this submanifold will not in general be the k-th jet of any subbundle of $\tilde{Y}$. Two distinct approaches are possible here. One could follow the *Lagrange-d'Alembert principle* and take variations of $\tilde{S}$ first, but choosing variations V (vertical vector fields on $\tilde{Y}$) such that the jet prolongations j^kV are tangent to the submanifold $G = 0$, and then enforce the constraint $G = 0$. On the other hand, one could consider the *variational nonholonomic* problem (also called *vakonomic*), and minimize $\tilde{S}$ over the set of all sections $(\tilde{\varphi}, \tilde{X})$ that satisfy the constraint $G = 0$, that is, enforce the constraint before taking the variations. If the constraint is holonomic, both approaches yield the same equations of motion. However, if the constraint is nonholonomic, the resulting equations are in general different. Which equations are correct is really a matter of experimental verification. It has been established that the Lagrange-d'Alembert principle gives the right equations of motion for nonholonomic mechanical systems, whereas the vakonomic setting is appropriate for optimal control problems (see [37], [38], [39], [40]).

We will argue that the vakonomic approach is the right one in our case. In Proposition 5 we showed that in the unconstrained case extremizing $S[\phi]$ with respect to ϕ was equivalent to extremizing $\tilde{S}[\tilde{\varphi}, \tilde{X}]$ with respect to $\tilde{\varphi}$, and in Proposition 6 we showed that extremizing with respect to $\tilde{X}$ did not yield new information. This is because there was no restriction on the fields $\tilde{\varphi}$ and $\tilde{X}$, and for any given $\tilde{X}$ there was a one-to-one correspondence between ϕ and $\tilde{\varphi}$ given by the formula $\varphi(t, x) = \phi(t, X(t, x))$, so extremizing over all possible $\tilde{\varphi}$ was equivalent to extremizing over all possible ϕ. Now, let $\mathcal{N}$ be the set of all smooth sections $(\tilde{\varphi}, \tilde{X})$ that satisfy the constraint (125) such that $X(t, .)$ is a diffeomorphism for all t. It should be intuitively clear that under appropriate assumptions on the mesh density function ρ, for any given smooth function $\phi(t, X)$, equation (25) together with $\varphi(t, x) = \phi(t, X(t, x))$ define a unique pair $(\tilde{\varphi}, \tilde{X}) \in \mathcal{N}$ (since our main purpose here is to only justify the application of the vakonomic approach, we do not attempt to specify those analytic assumptions precisely). Conversely, any given pair $(\tilde{\varphi}, \tilde{X}) \in \mathcal{N}$ defines a unique function ϕ through the formula $\phi(t, X) = \varphi(t, \xi(t, X))$, where $\xi(t, .) = X(t, .)^{-1}$, as in Section 3.1. Given this one-to-one correspondence and the fact that $S[\phi] = \tilde{S}[\tilde{\varphi}, \tilde{X}]$ by definition, we see that extremizing S with respect to all smooth ϕ is equivalent to extremizing $\tilde{S}$ over all smooth sections $(\tilde{\varphi}, \tilde{X}) \in \mathcal{N}$. We conclude that the vakonomic approach is appropriate in our case, since it follows from Hamilton's principle for the original, physically meaningful, action functional S.

Let us also note that our constraint depends on spatial derivatives only. Therefore, in the setting presented in Section 2 and Section 3 it can be considered holonomic, as it restricts the infinite-dimensional configuration manifold of fields that we used as our configuration space. In that case it is valid to use Hamilton's principle and minimize the action functional over the set of all allowable fields, i.e. those that satisfy the constraint $G = 0$. We did that by considering the augmented instantaneous Lagrangian (86).

In order to minimize $\tilde{S}$ over the set of sections satisfying the constraint (125) we will use the bundle-theoretic version of the Lagrange multiplier theorem, which we cite below after [41].

Theorem 1 (Lagrange multiplier theorem). *Let $\pi_{\mathcal{M},\mathcal{E}} : \mathcal{E} \longrightarrow \mathcal{M}$ be an inner product bundle over a smooth manifold $\mathcal{M}$, Ψ a smooth section of $\pi_{\mathcal{M},\mathcal{E}}$, and $h : \mathcal{M} \longrightarrow \mathbb{R}$ a smooth function. Setting $\mathcal{N} = \Psi^{-1}(0)$, the following are equivalent:*

1. *$\sigma \in \mathcal{N}$ is an extremum of $h|_{\mathcal{N}}$,*
2. *there exists an extremum $\bar{\sigma} \in \mathcal{E}$ of $\bar{h} : \mathcal{E} \longrightarrow \mathbb{R}$ such that $\pi_{\mathcal{M},\mathcal{E}}(\bar{\sigma}) = \sigma$,*

where $\bar{h}(\bar{\sigma}) = h(\pi_{\mathcal{M},\mathcal{E}}(\bar{\sigma})) - \langle \bar{\sigma}, \Psi(\pi_{\mathcal{M},\mathcal{E}}(\bar{\sigma})) \rangle_{\mathcal{E}}$.

Let us briefly review the ideas presented in [41], adjusting the notation to our problem and generalizing when necessary. Let

$$C^\infty_{\mathcal{U}}(\tilde{Y}) = \{\sigma = (\tilde{\varphi}, \tilde{X}) : \mathcal{U} \subset \mathcal{X} \longrightarrow \tilde{Y}\} \tag{126}$$

be the set of smooth sections of $\pi_{\mathcal{X}\tilde{Y}}$ on $\mathcal{U}$. Then $\tilde{S} : C^\infty_{\mathcal{U}}(\tilde{Y}) \longrightarrow \mathbb{R}$ can be identified with h in Theorem 1, where $\mathcal{M} = C^\infty_{\mathcal{U}}(\tilde{Y})$. Furthermore, define the trivial bundle

$$\pi_{\mathcal{X}\mathcal{V}} : \mathcal{V} = \mathcal{X} \times \mathbb{R} \longrightarrow \mathcal{X} \tag{127}$$

and let $C^\infty_{\mathcal{U}}(\mathcal{V})$ be the set of smooth sections $\tilde{\lambda} : \mathcal{U} \longrightarrow \mathcal{V}$, which represent our Lagrange multipliers and in local coordinates have the representation $\tilde{\lambda}(t, x) = (t, x, \lambda(t, x))$. The set $C^\infty_{\mathcal{U}}(\mathcal{V})$ is an inner product space with $\langle \tilde{\lambda}_1, \tilde{\lambda}_2 \rangle = \int_{\mathcal{U}} \lambda_1 \lambda_2 \, dt \wedge dx$. Take

$$\mathcal{E} = C^\infty_{\mathcal{U}}(\tilde{Y}) \times C^\infty_{\mathcal{U}}(\mathcal{V}). \tag{128}$$

This is an inner product bundle over $C^\infty_{\mathcal{U}}(\tilde{Y})$ with the inner product defined by

$$\Big\langle (\sigma, \tilde{\lambda}_1), (\sigma, \tilde{\lambda}_2) \Big\rangle_{\mathcal{E}} = \langle \tilde{\lambda}_1, \tilde{\lambda}_2 \rangle. \tag{129}$$

We now have to construct a smooth section $\Psi : C^\infty_{\mathcal{U}}(\tilde{Y}) \longrightarrow \mathcal{E}$ that will realize our constraint (125). Define the fiber-preserving mapping $\tilde{G} : J^k\tilde{Y} \longrightarrow \mathcal{V}$ such that for $\vartheta \in J^k\tilde{Y}$

$$\tilde{G}(\vartheta) = \big(\pi_{\mathcal{X}, J^k\tilde{Y}}(\vartheta), G(\vartheta)\big). \tag{130}$$

For instance, for $k = 1$, in local coordinates we have $\tilde{G}(t, x, y, v_t, v_x) = (t, x, G(t, x, y, v_t, v_x))$. Then we can define

$$\Psi(\sigma) = (\sigma, \tilde{G} \circ j^k\sigma). \tag{131}$$

The set of allowable sections $\mathcal{N} \subset C^\infty_{\mathcal{U}}(\tilde{Y})$ is now defined by $\mathcal{N} = \Psi^{-1}(0)$. That is, $(\tilde{\varphi}, \tilde{X}) \in \mathcal{N}$ provided that $G(j^k\tilde{\varphi}, j^k\tilde{X}) = 0$.

The augmented action functional $\tilde{S}_C : \mathcal{E} \longrightarrow \mathbb{R}$ is now given by

$$\tilde{S}_C[\bar{\sigma}] = \tilde{S}[\pi_{\mathcal{M},\mathcal{E}}(\bar{\sigma})] - \langle \bar{\sigma}, \Psi(\pi_{\mathcal{M},\mathcal{E}}(\bar{\sigma})) \rangle_{\mathcal{E}}, \tag{132}$$

or denoting $\bar{\sigma} = (\tilde{\varphi}, \tilde{X}, \tilde{\lambda})$

$$\begin{aligned}
\tilde{S}_C[\tilde{\varphi}, \tilde{X}, \tilde{\lambda}] &= \tilde{S}[\tilde{\varphi}, \tilde{X}] - \langle \tilde{\lambda}, \tilde{G} \circ (j^k\tilde{\varphi}, j^k\tilde{X}) \rangle \\
&= \int_{\mathcal{U}} \tilde{\mathcal{L}}(j^1\tilde{\varphi}, j^1\tilde{X}) \, dt \wedge dx - \int_{\mathcal{U}} \lambda(t, x) \, G(j^k\tilde{\varphi}, j^k\tilde{X}) \, dt \wedge dx \\
&= \int_{\mathcal{U}} \Big[\tilde{\mathcal{L}}(j^1\tilde{\varphi}, j^1\tilde{X}) - \lambda(t, x) \, G(j^k\tilde{\varphi}, j^k\tilde{X}) \Big] \, dt \wedge dx.
\end{aligned} \tag{133}$$

Theorem 1 states, that if $(\tilde{\varphi}, \tilde{X}, \tilde{\lambda})$ is an extremum of $\tilde{S}_C$, then $(\tilde{\varphi}, \tilde{X})$ extremizes $\tilde{S}$ over the set $\mathcal{N}$ of sections satisfying the constraint $G = 0$. Note that using the multisymplectic formalism we obtained the same result as (86) in the instantaneous formulation, where we could treat G as a holonomic constraint. The dynamics is obtained by solving for a triple $(\tilde{\varphi}, \tilde{X}, \tilde{\lambda})$ such that

$$\frac{d}{d\epsilon}\bigg|_{\epsilon=0} \tilde{S}_C[\eta^\epsilon_Y \circ \tilde{\varphi}, \eta^\epsilon_{\mathcal{B}} \circ \tilde{X}, \eta^\epsilon_{\mathcal{V}} \circ \tilde{\lambda}] = 0 \tag{134}$$

for all η^ϵ_Y, $\eta^\epsilon_{\mathcal{B}}$, $\eta^\epsilon_{\mathcal{V}}$ that keep the boundary conditions on $\partial\mathcal{U}$ fixed, where η^ϵ denotes the flow of vertical vector fields on respective bundles.

Note that we can define $\tilde{Y}_C = Y \oplus \mathcal{B} \oplus \mathcal{V}$ and $\tilde{\mathcal{L}_C} : J^k\tilde{Y}_C \longrightarrow \mathbb{R}$ by setting $\tilde{\mathcal{L}_C} = \tilde{\mathcal{L}} - \lambda \cdot G$, i.e., we can consider a k-th order field theory. If $k = 1, 2$ then an appropriate multisymplectic form formula in terms of the fields $\tilde{\varphi}$, $\tilde{X}$ and $\tilde{\lambda}$ will hold. Presumably, this can be generalized for $k > 2$ using the techniques put forth in [35]. However, it is an interesting question whether there exists any multisymplectic form formula defined in terms of $\tilde{\varphi}$, $\tilde{X}$ and objects on $J^k\tilde{Y}$ only. It appears to be an open problem. This would be the multisymplectic analog of the fact that the flow of a constrained mechanical system is symplectic on the constraint submanifold of the configuration space.

Discretization

Let us use the same discretization as discussed in Section 4.2. Assume we have a discrete Lagrangian $\tilde{L} : J^1\tilde{Y} \longrightarrow \mathbb{R}$, the corresponding discrete action $\tilde{S}[\tilde{\varphi}, \tilde{X}]$, and a discrete constraint $G : J^1\tilde{Y} \longrightarrow \mathbb{R}$ or $G : J^2_0\tilde{Y} \longrightarrow \mathbb{R}$. Note that $\tilde{S}$ is essentially a function of $2MN$ variables and we want to extremize it subject to the set of algebraic constraints $G = 0$. The standard Lagrange multiplier theorem proved in basic calculus textbooks applies here. However, let us work out a discrete counterpart of the formalism introduced at the continuous level. This will facilitate the discussion of the discrete notion of multisymplecticity. Let

$$C_{\mathcal{U}}(\tilde{Y}) = \{\sigma = (\tilde{\varphi}, \tilde{X}) : \mathcal{U} \subset \mathcal{X} \longrightarrow \tilde{Y}\} \tag{135}$$

be the set of discrete sections of $\pi_{\mathcal{X}\tilde{Y}} : \tilde{Y} \longrightarrow \mathcal{X}$. Similarly, define the discrete bundle $\mathcal{V} = \mathcal{X} \times \mathbb{R}$ and let $C_{\mathcal{U}_0}(\mathcal{V})$ be the set of discrete sections $\tilde{\lambda} : \mathcal{U}_0 \longrightarrow \mathcal{V}$ representing the Lagrange multipliers, where $\mathcal{U}_0 \subset \mathcal{U}$ is defined below. Let $\tilde{\lambda}(j, i) = (j, i, \lambda(j, i))$ with $\lambda^j_i \equiv \lambda(j, i)$ be the local representation. The set $C_{\mathcal{U}_0}(\mathcal{V})$ is an inner product space with $\langle \tilde{\lambda}, \tilde{\mu} \rangle = \sum_{(j,i)\in\mathcal{U}_0} \lambda^j_i \mu^j_i$. Take $\mathcal{E} = C_{\mathcal{U}}(\tilde{Y}) \times C_{\mathcal{U}_0}(\mathcal{V})$. Just like at the continuous level, $\mathcal{E}$ is an inner product bundle. However, at the discrete level it is more convenient to define the inner product on $\mathcal{E}$ in a slightly modified way. Since there are some nuances in the notation, let us consider the cases $k = 1$ and $k = 2$ separately.

Case $k = 1$.

Let $\mathcal{U}_0 = \{(j, i) \in \mathcal{U} \,|\, j \leq M, i \leq N\}$. Define the trivial bundle $\hat{\mathcal{V}} = \mathcal{X}^\Box \times \mathbb{R}$ and let $C_{\mathcal{U}^\Box}(\hat{\mathcal{V}})$ be the set of all sections of $\hat{\mathcal{V}}$ defined on $\mathcal{U}^\Box$. For a given section $\tilde{\lambda} \in C_{\mathcal{U}_0}(\mathcal{V})$ we define its extension $\hat{\lambda} \in C_{\mathcal{U}^\Box}(\hat{\mathcal{V}})$ by

$$\hat{\lambda}(\Box) = \big(\Box, \lambda(\Box^1)\big), \tag{136}$$

that is, $\hat{\lambda}$ assigns to the square $\Box$ the value that $\tilde{\lambda}$ takes on the *first* vertex of that square. Note that this operation is invertible: given a section of $C_{\mathcal{U}^\Box}(\hat{\mathcal{V}})$ we can uniquely determine a section of $C_{\mathcal{U}_0}(\mathcal{V})$. We can define the inner product

$$\langle \hat{\lambda}, \hat{\mu} \rangle = \sum_{\Box \subset \mathcal{U}} \lambda(\Box^1)\mu(\Box^1). \tag{137}$$

One can easily see that we have $\langle \hat{\lambda}, \hat{\mu} \rangle = \langle \tilde{\lambda}, \tilde{\mu} \rangle$, so by a slight abuse of notation we can use the same symbol $\langle ., . \rangle$ for both inner products. It will be clear from the context which definition should be invoked. We can now define an inner product on the fibers of $\mathcal{E}$ as

$$\Big\langle (\sigma, \tilde{\lambda}), (\sigma, \tilde{\mu}) \Big\rangle_{\mathcal{E}} = \langle \hat{\lambda}, \hat{\mu} \rangle = \langle \tilde{\lambda}, \tilde{\mu} \rangle. \tag{138}$$

Let us now construct a section $\Psi : C_{\mathcal{U}}(\tilde{Y}) \longrightarrow \mathcal{E}$ that will realize our discrete constraint G. First, in analogy to (130), define the fiber-preserving mapping $\tilde{G} : J^1\tilde{Y} \longrightarrow \hat{\mathcal{V}}$ such that

$$\tilde{G}(y_{\Box^l}, X_{\Box^r}) = \big(\Box, G(y_{\Box^l}, X_{\Box^r})\big), \tag{139}$$

where $l, r = 1, 2, 3, 4$. We now define Ψ by requiring that for $\sigma \in C_{\mathcal{U}}(\tilde{Y})$ the extension (136) of $\Psi(\sigma)$ is given by

$$\hat{\Psi}(\sigma) = (\sigma, \tilde{G} \circ j^1\sigma). \tag{140}$$

The set of allowable sections $\mathcal{N} \subset C_{\mathcal{U}}(\tilde{Y})$ is now defined by $\mathcal{N} = \Psi^{-1}(0)$—that is, $(\tilde{\varphi}, \tilde{X}) \in \mathcal{N}$ provided that $G(j^1\tilde{\varphi}, j^1\tilde{X}) = 0$ for all $\Box \in \mathcal{U}^{\Box}$. The augmented discrete action $\tilde{S}_C : \mathcal{E} \longrightarrow \mathbb{R}$ is therefore

$$\begin{aligned}
\tilde{S}_C[\sigma, \tilde{\lambda}] &= \tilde{S}[\sigma] - \Big\langle (\sigma, \tilde{\lambda}), \Psi(\sigma) \Big\rangle_{\mathcal{E}} \\
&= \tilde{S}[\sigma] - \Big\langle \hat{\lambda}, \tilde{G} \circ j^1\sigma \Big\rangle \\
&= \sum_{\Box \subset \mathcal{U}} \tilde{L}(j^1\sigma) - \sum_{\Box \subset \mathcal{U}} \lambda(\Box^1) G(j^1\sigma) \\
&= \sum_{\Box \subset \mathcal{U}} \Big(\tilde{L}(j^1\sigma) - \lambda(\Box^1) G(j^1\sigma) \Big).
\end{aligned} \tag{141}$$

By the standard Lagrange multiplier theorem, if $(\tilde{\varphi}, \tilde{X}, \tilde{\lambda})$ is an extremum of $\tilde{S}_C$, then $(\tilde{\varphi}, \tilde{X})$ is an extremum of $\tilde{S}$ over the set $\mathcal{N}$ of sections satisfying the constraint $G = 0$. The discrete Hamilton principle can be expressed as

$$\frac{d}{d\epsilon}\bigg|_{\epsilon=0} \tilde{S}_C[\tilde{\varphi}_\epsilon, \tilde{X}_\epsilon, \tilde{\lambda}_\epsilon] = 0 \tag{142}$$

for all vector fields V on Y, W on $\mathcal{B}$, and Z on $\mathcal{V}$ that keep the boundary conditions on $\partial\mathcal{U}$ fixed, where $\tilde{\varphi}_\epsilon(j,i) = F_\epsilon^{V_{ji}}(\tilde{\varphi}(j,i))$ and $F_\epsilon^{V_{ji}}$ is the flow of V_{ji} on $\mathbb{R}$, and similarly for $\tilde{X}_\epsilon$ and $\tilde{\lambda}_\epsilon$. The discrete Euler-Lagrange equations can be conveniently computed if in (142) one focuses on some $(j,i) \in \text{int}\,\mathcal{U}$. With the convention $\tilde{\varphi}(j,i) = y_i^j$, $\tilde{X}(j,i) = X_i^j$, $\tilde{\lambda}(j,i) = \lambda_i^j$, we write the terms of $\tilde{S}_C$ containing y_i^j, X_i^j and λ_i^j explicitly as

$$\begin{aligned}
\tilde{S}_C = \ldots &+ \tilde{L}\big(y_i^j, y_{i+1}^j, y_{i+1}^{j+1}, y_i^{j+1}, X_i^j, X_{i+1}^j, X_{i+1}^{j+1}, X_i^{j+1}\big) \\
&+ \tilde{L}\big(y_{i-1}^j, y_i^j, y_i^{j+1}, y_{i-1}^{j+1}, X_{i-1}^j, X_i^j, X_i^{j+1}, X_{i-1}^{j+1}\big) \\
&+ \tilde{L}\big(y_{i-1}^{j-1}, y_i^{j-1}, y_i^j, y_{i-1}^j, X_{i-1}^{j-1}, X_i^{j-1}, X_i^j, X_{i-1}^j\big) \\
&+ \tilde{L}\big(y_i^{j-1}, y_{i+1}^{j-1}, y_{i+1}^j, y_i^j, X_i^{j-1}, X_{i+1}^{j-1}, X_{i+1}^j, X_i^j\big) \\
&+ \lambda_i^j G\big(y_i^j, y_{i+1}^j, y_{i+1}^{j+1}, y_i^{j+1}, X_i^j, X_{i+1}^j, X_{i+1}^{j+1}, X_i^{j+1}\big) \\
&+ \lambda_{i-1}^j G\big(y_{i-1}^j, y_i^j, y_i^{j+1}, y_{i-1}^{j+1}, X_{i-1}^j, X_i^j, X_i^{j+1}, X_{i-1}^{j+1}\big) \\
&+ \lambda_{i-1}^{j-1} G\big(y_{i-1}^{j-1}, y_i^{j-1}, y_i^j, y_{i-1}^j, X_{i-1}^{j-1}, X_i^{j-1}, X_i^j, X_{i-1}^j\big) \\
&+ \lambda_i^{j-1} G\big(y_i^{j-1}, y_{i+1}^{j-1}, y_{i+1}^j, y_i^j, X_i^{j-1}, X_{i+1}^{j-1}, X_{i+1}^j, X_i^j\big) + \ldots
\end{aligned} \tag{143}$$

The discrete Euler-Lagrange equations are obtained by differentiating with respect to y_i^j, X_i^j and λ_i^j, and can be written compactly as

$$\sum_{\substack{l,\square \\ (j,i)=\square^l}} \left[\frac{\partial \tilde{L}}{\partial y^l}(y_{\square^1},\dots,y_{\square^4},X_{\square^1},\dots,X_{\square^4}) + \right.$$

$$\left. + \lambda_{\square^1} \frac{\partial G}{\partial y^l}(y_{\square^1},\dots,y_{\square^4},X_{\square^1},\dots,X_{\square^4}) \right] = 0,$$

$$\sum_{\substack{l,\square \\ (j,i)=\square^l}} \left[\frac{\partial \tilde{L}}{\partial X^l}(y_{\square^1},\dots,y_{\square^4},X_{\square^1},\dots,X_{\square^4}) + \right.$$

$$\left. + \lambda_{\square^1} \frac{\partial G}{\partial X^l}(y_{\square^1},\dots,y_{\square^4},X_{\square^1},\dots,X_{\square^4}) \right] = 0,$$

$$G(y_i^j, y_{i+1}^j, y_{i+1}^{j+1}, y_i^{j+1}, X_i^j, X_{i+1}^j, X_{i+1}^{j+1}, X_i^{j+1}) = 0 \tag{144}$$

for all $(j,i) \in \mathrm{int}\,\mathcal{U}$. If we know y_i^{j-1}, X_i^{j-1}, y_i^j, X_i^j and λ_i^{j-1} for $i = 1,\dots,N$, this system of equations allows us to solve for y_i^{j+1}, X_i^{j+1} and λ_i^j.

Note that we can define $\tilde{Y}_C = Y \oplus \mathcal{B} \oplus \mathcal{V}$ and the augmented Lagrangian $\tilde{L}_C : J^1\tilde{Y}_C \longrightarrow \mathbb{R}$ by setting

$$\tilde{L}_C(j^1\tilde{\varphi}, j^1\tilde{X}, j^1\tilde{\lambda}) = \tilde{L}(j^1\tilde{\varphi}, j^1\tilde{X}) - \lambda(\square^1) \cdot G(j^1\tilde{\varphi}, j^1\tilde{X}), \tag{145}$$

that is, we can consider an unconstrained field theory in terms of the fields $\tilde{\varphi}$, $\tilde{X}$ and $\tilde{\lambda}$. Then, the solutions of (144) satisfy the multisymplectic form formula (110) in terms of objects defined on $J^1\tilde{Y}_C$.

Case $k = 2$.

Let $\mathcal{U}_0 = \{(j,i) \in \mathcal{U} \,|\, j \le M, 1 \le i \le N\}$. Define the trivial bundle $\hat{\mathcal{V}} = \mathcal{X}^{\boxplus} \times \mathbb{R}$ and let $C_{\mathcal{U}^{\boxplus}}(\hat{\mathcal{V}})$ be the set of all sections of $\hat{\mathcal{V}}$ defined on $\mathcal{U}^{\boxplus}$. For a given section $\tilde{\lambda} \in C_{\mathcal{U}_0}(\mathcal{V})$ we define its extension $\hat{\lambda} \in C_{\mathcal{U}^{\boxplus}}(\hat{\mathcal{V}})$ by

$$\hat{\lambda}(\boxplus) = \big(\boxplus, \lambda(\boxplus^2)\big), \tag{146}$$

that is, $\hat{\lambda}$ assigns to the 6-tuple $\boxplus$ the value that $\tilde{\lambda}$ takes on the *second* vertex of that 6-tuple. Like before, this operation is invertible. We can define the inner product

$$\langle \hat{\lambda}, \hat{\mu} \rangle = \sum_{\boxplus \subset \mathcal{U}} \lambda(\boxplus^2)\mu(\boxplus^2) \tag{147}$$

and the inner product on $\mathcal{E}$ as in (138). Define the fiber-preserving mapping $\tilde{G} : J_0^2\tilde{Y} \longrightarrow \hat{\mathcal{V}}$ such that

$$\tilde{G}(y_{\boxplus^l}, X_{\boxplus^r}) = \big(\boxplus, G(y_{\boxplus^l}, X_{\boxplus^r})\big), \tag{148}$$

where $l, r = 1,\dots,6$. We now define Ψ by requiring that for $\sigma \in C_{\mathcal{U}}(\tilde{Y})$ the extension (146) of $\Psi(\sigma)$ is given by

$$\hat{\Psi}(\sigma) = (\sigma, \tilde{G} \circ j_0^2\sigma). \tag{149}$$

Again, the set of allowable sections is $\mathcal{N} = \Psi^{-1}(0)$. That is, $(\tilde{\varphi}, \tilde{X}) \in \mathcal{N}$ provided that $G(j_0^2\tilde{\varphi}, j_0^2\tilde{X}) = 0$ for all $\boxplus \in \mathcal{U}^{\boxplus}$. The augmented discrete action $\tilde{S}_C : \mathcal{E} \longrightarrow \mathbb{R}$ is therefore

$$\begin{aligned}\tilde{S}_C[\sigma,\tilde{\lambda}] &= \tilde{S}[\sigma] - \Big\langle (\sigma,\tilde{\lambda}), \Psi(\sigma)\Big\rangle_{\mathcal{E}} \\ &= \tilde{S}[\sigma] - \Big\langle \hat{\lambda}, \tilde{G}\circ j_0^2\sigma \Big\rangle \\ &= \sum_{\square\subset\mathcal{U}} \tilde{L}(j^1\sigma) - \sum_{\boxplus\subset\mathcal{U}} \lambda(\boxplus^2)G(j_0^2\sigma).\end{aligned} \tag{150}$$

Writing out the terms involving y_i^j, X_i^j and λ_i^j explicitly, as in (143), and invoking the discrete Hamilton principle (142), one obtains the discrete Euler-Lagrange equations, which can be compactly expressed as

$$\sum_{\substack{l,\square \\ (j,i)=\square^l}} \frac{\partial \tilde{L}}{\partial y^l}(y_{\square^1},\dots,y_{\square^4},X_{\square^1},\dots,X_{\square^4}) + \\ + \sum_{\substack{l,\boxplus \\ (j,i)=\boxplus^l}} \lambda_{\boxplus^2}\frac{\partial G}{\partial y^l}(y_{\boxplus^1},\dots,y_{\boxplus^6},X_{\boxplus^1},\dots,X_{\boxplus^6}) = 0,$$

$$\sum_{\substack{l,\square \\ (j,i)=\square^l}} \frac{\partial \tilde{L}}{\partial X^l}(y_{\square^1},\dots,y_{\square^4},X_{\square^1},\dots,X_{\square^4}) + \\ + \sum_{\substack{l,\boxplus \\ (j,i)=\boxplus^l}} \lambda_{\boxplus^2}\frac{\partial G}{\partial X^l}(y_{\boxplus^1},\dots,y_{\boxplus^6},X_{\boxplus^1},\dots,X_{\boxplus^6}) = 0,$$

$$G(y_{i-1}^j,y_i^j,y_{i+1}^j,y_{i+1}^{j+1},y_i^{j+1},y_{i-1}^{j+1},X_{i-1}^j,X_i^j,X_{i+1}^j,X_{i+1}^{j+1},X_i^{j+1},X_{i-1}^{j+1}) = 0 \tag{151}$$

for all $(j,i)\in \mathrm{int}\,\mathcal{U}$. If we know y_i^{j-1}, X_i^{j-1}, y_i^j, X_i^j and λ_i^{j-1} for $i=1,\dots,N$, this system of equations allows us to solve for y_i^{j+1}, X_i^{j+1} and λ_i^j.

Let us define the extension $\tilde{L}_{\text{ext}}: J_0^2\tilde{Y} \longrightarrow \mathbb{R}$ of the Lagrangian density $\tilde{L}$ by setting

$$\tilde{L}_{\text{ext}}(y_{\boxplus^1},\dots,X_{\boxplus^6}) = \begin{cases} \tilde{L}(y_{\square^1},\dots,X_{\square^4}) & \text{if } \boxplus^2=(j,0),(j,N+1), \\ & \text{where } \square=\boxplus\cap\mathcal{U}, \\ \frac{1}{2}\sum_{\square\subset\boxplus}\tilde{L}(y_{\square^1},\dots,X_{\square^4}) & \text{otherwise.}\end{cases} \tag{152}$$

Let us also set $G(y_{\square^1},\dots,X_{\square^4})=0$ if $\boxplus^2=(j,0),(j,N+1)$. Define $\mathcal{A}=\{\boxplus \mid \boxplus^2,\boxplus^5\in\mathcal{U}\}$. Then (150) can be written as

$$\tilde{S}_C[\sigma,\tilde{\lambda}] = \sum_{\boxplus\in\mathcal{A}}\Big[\tilde{L}_{\text{ext}}(j_0^2\sigma) - \lambda(\boxplus^2)G(j_0^2\sigma)\Big] = \sum_{\boxplus\in\mathcal{A}}\tilde{L}_C(j_0^2\sigma, j_0^2\tilde{\lambda}), \tag{153}$$

where the last equality defines the augmented Lagrangian $\tilde{L}_C: J_0^2\tilde{Y}_C \longrightarrow \mathbb{R}$ for $\tilde{Y}_C = Y\oplus\mathcal{B}\oplus\mathcal{V}$. Therefore, we can consider an unconstrained second-order field theory in terms of the fields $\tilde{\varphi}$, $\tilde{X}$ and $\tilde{\lambda}$, and the solutions of (151) will satisfy a discrete multisymplectic form formula very similar to the one proved in [35]. The only difference is the fact that the authors analyzed a discretization of the Camassa-Holm equation and were able to consider an even smaller subbundle of the second jet of the configuration bundle. As a result it was sufficient for them to consider a discretization based on squares □ rather than 6-tuples ◫. In our case there will be six discrete 2-forms $\Omega^l_{\tilde{L}_C}$ for $l = 1, \ldots, 6$ instead of just four.

Remark.

In both cases we showed that our discretization leads to integrators that are multisymplectic on the augmented jets $J^k\tilde{Y}_C$. However, just like in the continuous setting, it is an interesting problem whether there exists a discrete multisymplectic form formula in terms of objects defined on $J^k\tilde{Y}$ only.

Example: Trapezoidal rule.

Consider the semi-discrete Lagrangian (44). We can use the trapezoidal rule to define the discrete Lagrangian (14) as

$$\tilde{L}_d(y^j, X^j, y^{j+1}, X^{j+1}) = \frac{\Delta t}{2}\tilde{L}_N\Big(y^j, X^j, \frac{y^{j+1}-y^j}{\Delta t}, \frac{X^{j+1}-X^j}{\Delta t}\Big) + \frac{\Delta t}{2}\tilde{L}_N\Big(y^{j+1}, X^{j+1}, \frac{y^{j+1}-y^j}{\Delta t}, \frac{X^{j+1}-X^j}{\Delta t}\Big), \tag{154}$$

where $y^j = (y^j_1, \ldots, y^j_N)$ and $X^j = (X^j_1, \ldots, X^j_N)$. The constrained version (see [2]) of the Discrete Euler-Lagrange equations (103) takes the form

$$\begin{aligned} D_2\tilde{L}_d(q^{j-1}, q^j) + D_1\tilde{L}_d(q^j, q^{j+1}) &= Dg(q^j)^T\lambda^j, \\ g(q^{j+1}) &= 0, \end{aligned} \tag{155}$$

where for brevity $q^j = (y^j_1, X^j_1, \ldots, y^j_N, X^j_N)$, $\lambda^j = (\lambda^j_1, \ldots, \lambda^j_N)$ and g is an adaptation constraint, for instance (28). If q^{j-1}, q^j are known, then (155) can be used to compute q^{j+1} and λ^j. It is easy to verify that the condition (94) is enough to ensure solvability of (155), assuming the time step Δt is sufficiently small, so there is no need to introduce slack degrees of freedom as in (95). If the mass matrix (47) was constant and nonsingular, then (155) would result in the SHAKE algorithm, or in the RATTLE algorithm if one passes to the position-momentum formulation (see [1], [2]).

Using (38) and (41) we can write

$$\tilde{L}_d(y^j, X^j, y^{j+1}, X^{j+1}) = \sum_{i=0}^{N}\tilde{L}(y^j_i, y^j_{i+1}, y^{j+1}_{i+1}, y^{j+1}_i, X^j_i, X^j_{i+1}, X^{j+1}_{i+1}, X^{j+1}_i), \tag{156}$$

where we defined the discrete Lagrangian $\tilde{L} : J^1\tilde{Y} \longrightarrow \mathbb{R}$ by the formula

$$\begin{aligned} &\tilde{L}\Big(y^j_i, y^j_{i+1}, y^{j+1}_{i+1}, y^{j+1}_i, X^j_i, X^j_{i+1}, X^{j+1}_{i+1}, X^{j+1}_i\Big) = \\ &\quad \frac{\Delta t}{2}\int_{x_i}^{x_{i+1}} \tilde{\mathcal{L}}\Big(\bar{\varphi}^j(x), \bar{X}^j(x), \bar{\varphi}^j_x(x), \bar{X}^j_x(x), \bar{\varphi}_t(x), \bar{X}_t(x)\Big)\,dx \\ &\qquad + \frac{\Delta t}{2}\int_{x_i}^{x_{i+1}} \tilde{\mathcal{L}}\Big(\bar{\varphi}^{j+1}(x), \bar{X}^{j+1}(x), \bar{\varphi}^{j+1}_x(x), \bar{X}^{j+1}_x(x), \bar{\varphi}_t(x), \bar{X}_t(x)\Big)\,dx \end{aligned} \tag{157}$$

with

$$
\begin{aligned}
\bar{\varphi}^j(x) &= y_i^j \eta_i(x) + y_{i+1}^j \eta_{i+1}(x), \\
\bar{\varphi}_x^j(x) &= \frac{y_{i+1}^j - y_i^j}{\Delta x}, \\
\bar{\varphi}_t(x) &= \frac{y_i^{j+1} - y_i^j}{\Delta t}\eta_i(x) + \frac{y_{i+1}^{j+1} - y_{i+1}^j}{\Delta t}\eta_{i+1}(x),
\end{aligned} \tag{158}
$$

and similarly for $\bar{X}(x)$. Given the Lagrangian density $\tilde{\mathcal{L}}$ as in (43) one can compute the integrals in (157) explicitly. Suppose that the adaptation constraint g has a 'local' structure, for instance

$$
g_i(y^j, X^j) = G(y_i^j, y_{i+1}^j, y_{i+1}^{j+1}, y_i^{j+1}, X_i^j, X_{i+1}^j, X_{i+1}^{j+1}, X_i^{j+1}), \tag{159}
$$

as in (122) or

$$
g_i(y^j, X^j) = G(y_{\boxplus^l}, X_{\boxplus^r}), \qquad \text{where } \boxplus^2 = (j,i), \tag{160}
$$

as in (123). It is straightforward to show that (144) or (151) are equivalent to (155), that is, the variational integrator defined by (155) is also multisymplectic.

For reasons similar to the ones pointed out in Section 4.2, the 2-nd and 4-th order Lobatto IIIA-IIIB methods that we used for our numerical computations are not multisymplectic.

5. Numerical results

5.1. The Sine-Gordon equation

We applied the methods discussed in the previous sections to the Sine-Gordon equation

$$
\frac{\partial^2 \phi}{\partial t^2} - \frac{\partial^2 \phi}{\partial X^2} + \sin\phi = 0. \tag{161}
$$

This equation results from the (1+1)-dimensional scalar field theory with the Lagrangian density

$$
\mathcal{L}(\phi, \phi_X, \phi_t) = \frac{1}{2}\phi_t^2 - \frac{1}{2}\phi_X^2 - (1 - \cos\phi). \tag{162}
$$

The Sine-Gordon equation arises in many physical applications. For instance, it governs the propagation of dislocations in crystals, the evolution of magnetic flux in a long Josephson-junction transmission line or the modulation of a weakly unstable baroclinic wave packet in a two-layer fluid. It also has applications in the description of one-dimensional organic conductors, one-dimensional ferromagnets, liquid crystals, or in particle physics as a model for baryons (see [42], [43]).

The Sine-Gordon equation has interesting soliton solutions. A single soliton traveling at the speed v is given by

$$
\phi_S(X,t) = 4\arctan\left[\exp\left(\frac{X - X_0 - vt}{\sqrt{1-v^2}}\right)\right]. \tag{163}
$$

It is depicted in Figure 2. The backscattering of two solitons, each traveling with the velocity v, is described by the formula

$$
\phi_{SS}(X,t) = 4\arctan\left[\frac{v\sinh(\frac{X}{\sqrt{1-v^2}})}{\cosh(\frac{vt}{\sqrt{1-v^2}})}\right]. \tag{164}
$$

It is depicted in Figure 3. Note that if we restrict $X \geq 0$, then this formula also gives a single soliton solution satisfying the boundary condition $\phi(0,t) = 0$, that is, a soliton bouncing from a rigid wall.

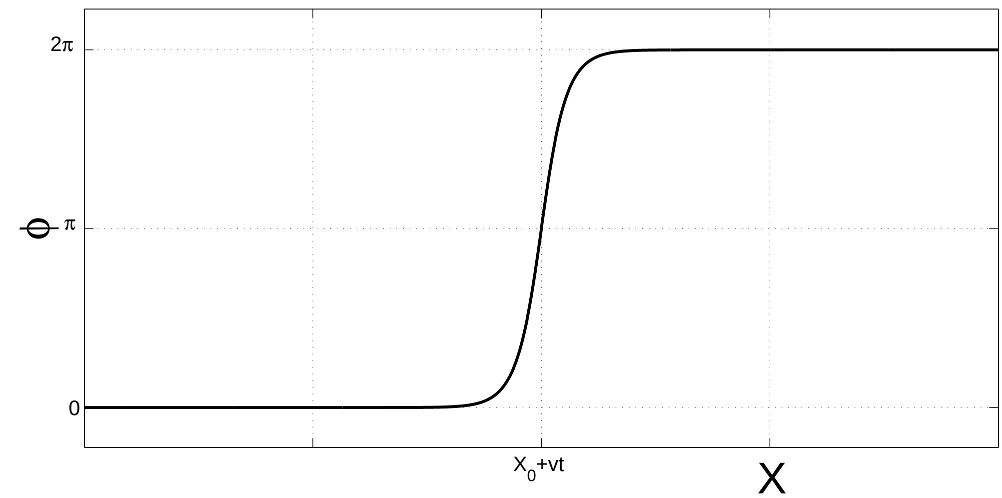

Figure 2. The single soliton solution of the Sine-Gordon equation.

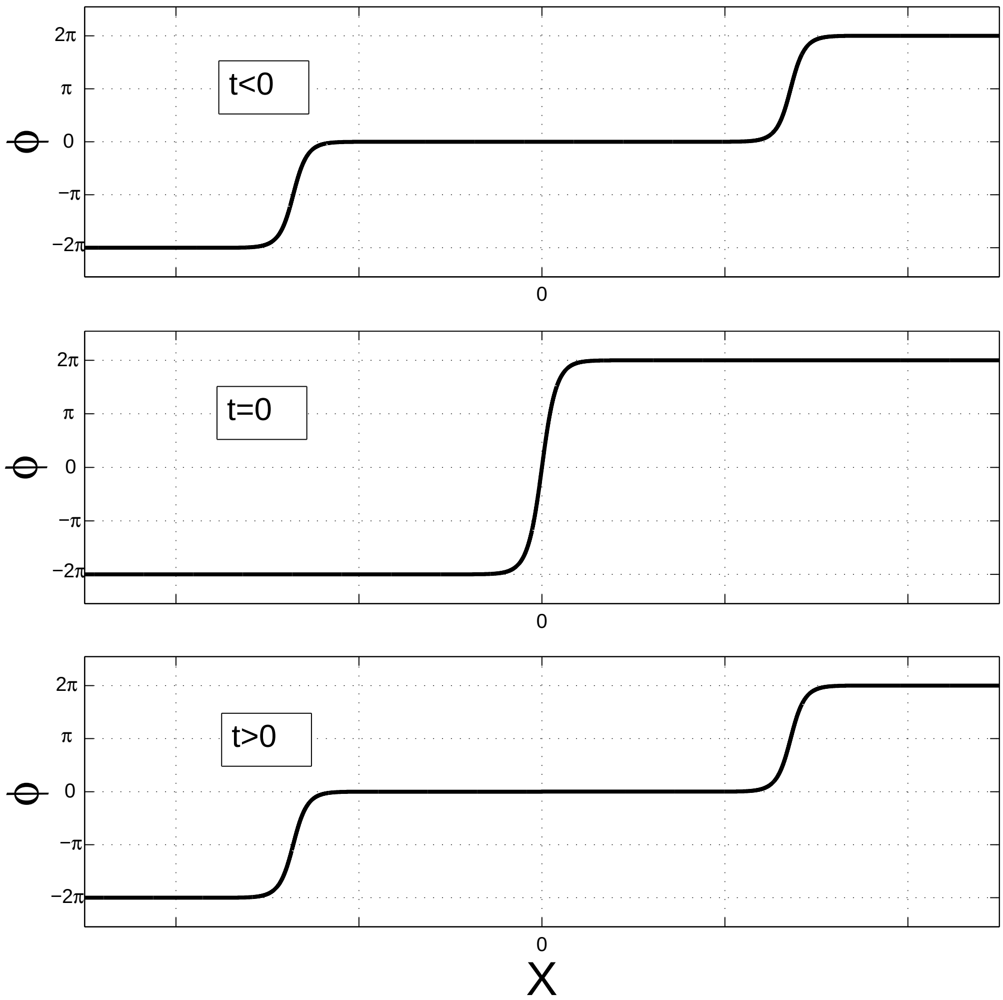

Figure 3. The two-soliton solution of the Sine-Gordon equation.

5.2. Generating consistent initial conditions

Suppose we specify the following initial conditions

$$\begin{aligned} \phi(X,0) &= a(X), \\ \phi_t(X,0) &= b(X), \end{aligned} \tag{165}$$

and assume they are consistent with the boundary conditions (2). In order to determine appropriate consistent initial conditions for (18) and (98) we need to solve several equations. First we solve for the y_i's and X_i's. We have $y_0 = \phi_L$, $y_{N+1} = \phi_R$, $X_0 = 0$, $X_{N+1} = X_{max}$. The rest are determined by solving the system

$$\begin{aligned} y_i &= a(X_i), \\ 0 &= g_i(y_1,\dots,y_N,X_1,\dots,X_N), \end{aligned} \tag{166}$$

for $i = 1,\dots,N$. This is a system of $2N$ nonlinear equations for $2N$ unknowns. We solve it using Newton's method. Note, however, that we do not *a priori* know good starting points for Newton's iterations. If our initial guesses are not close enough to the desired solution, the iterations may converge to the wrong solution or may not converge at all. In our computations we used the constraints (28). We found that a very simple variant of a homotopy continuation method worked very well in our case. Note that for $\alpha = 0$ the set of constraints (28) generates a uniform mesh. In order to solve (166) for some $\alpha > 0$, we split $[0,\alpha]$ into d subintervals by picking $\alpha_k = (k/d)\cdot\alpha$ for $k = 1,\dots,d$. We then solved (166) with α_1 using the uniformly spaced mesh points $X_i^{(0)} = (i/(N+1))\cdot X_{max}$ as our initial guess, resulting in $X_i^{(1)}$ and $y_i^{(1)}$. Then we solved (166) with α_2 using $X_i^{(1)}$ and $y_i^{(1)}$ as the initial guesses, resulting in $X_i^{(2)}$ and $y_i^{(2)}$. Continuing in this fashion, we got $X_i^{(d)}$ and $y_i^{(d)}$ as the numerical solution to (166) for the original value of α. Note that for more complicated initial conditions and constraint functions, predictor-corrector methods should be used—see [44] for more information. Another approach to solving (166) could be based on relaxation methods (see [7], [8]).

Next, we solve for the initial values of the velocities $\dot{y}_i$ and $\dot{X}_i$. Since $\varphi(x,t) = \phi(X(x,t),t)$, we have $\varphi_t(x,t) = \phi_X(X(x,t),t)X_t(x,t) + \phi_t(X(x,t),t)$. We also require that the velocities be consistent with the constraints. Hence the linear system

$$\begin{aligned} \dot{y}_i &= a'(X_i)\dot{X}_i + b(X_i), \qquad\qquad i = 1,\dots,N \\ 0 &= \frac{\partial g}{\partial y}(y,X)\dot{y} + \frac{\partial g}{\partial X}(y,X)\dot{X}. \end{aligned} \tag{167}$$

This is a system of $2N$ linear equations for the $2N$ unknowns $\dot{y}_i$ and $\dot{X}_i$, where $y = (y_1,\dots,y_N)$ and $X = (X_1,\dots,X_N)$. We can use those velocities to compute the initial values of the conjugate momenta. For the control-theoretic approach we use $p_i = \partial\tilde{L}_N/\partial\dot{y}_i$, as in Section 2.3, and for the Lagrange multiplier approach we use (46). In addition, for the Lagrange multiplier approach we also have the initial values for the slack variables $r_i = 0$ and their conjugate momenta $B_i = \partial\tilde{L}_N^A/\partial\dot{r}_i = 0$. It is also useful to use (93) to compute the initial values of the Lagrange multipliers λ_i that can be used as initial guesses in the first iteration of the Lobatto IIIA-IIIB algorithm. The initial guesses for the slack Lagrange multipliers are trivially $\mu_i = 0$. Note that both λ and μ are algebraic variables, so their values at each time step are completely determined by the Lobatto IIIA-IIIB algorithm (see [1], [27], [28] for details), and therefore no further initial or boundary conditions are necessary.

5.3. Convergence

In order to test the convergence of our methods as the number of mesh points N is increased, we considered a single soliton bouncing from two rigid walls at $X = 0$ and $X = X_{max} = 25$. We imposed the boundary conditions $\phi_L = 0$ and $\phi_R = 2\pi$, and as initial conditions we used (163) with $X_0 = 12.5$ and $v = 0.9$. It is possible to obtain the exact solution to this problem by considering a multi-soliton solution to (161) on the whole real line. Such a solution can be obtained using a Bäcklund transformation (see [42], [43]). However, the formulas quickly become complicated and, technically, one would have to consider an infinite number of solitons. Instead, we constructed a nearly exact solution by approximating the boundary interactions with (164):

$$\phi_{exact}(X,t) = \begin{cases} \phi_{SS}\big(X - X_{max}, t - (4n+1)T\big) + 2\pi & \text{if } t \in \big[4nT, (4n+2)T\big), \\ \phi_{SS}\big(X, t - (4n+3)T\big) & \text{if } t \in \big[(4n+2)T, (4n+4)T\big), \end{cases} \tag{168}$$

where n is an integer number and T satisfies $\phi_{SS}(X_{max}/2, T) = \pi$ (we numerically found $T \approx 13.84$). Given how fast (163) and (164) approach its asymptotic values, one may check that (168) can be considered exact to machine precision.

We performed numerical integration with the constant time step $\Delta t = 0.01$ up to the time $T_{max} = 50$. For the control-theoretic strategy we used the 1-stage and 2-stage Gauss method (2-nd and 4-th order respectively), and the 2-stage and 3-stage Lobatto IIIA-IIIB method (also 2-nd/4-th order). For the Lagrange multiplier strategy we used the 2-stage and 3-stage Lobatto IIIA-IIIB method for constrained mechanical systems (2-nd/4-th order). See [1], [14], [12] for more information about the mentioned symplectic Runge-Kutta methods. We used the constraints (28) based on the generalized arclength density (26). We chose the scaling parameter to be $\alpha = 2.5$, so that approximately half of the available mesh points were concentrated in the area of high gradient. A few example solutions are presented in Figure 4-7. Note that the Lagrange multiplier strategy was able to accurately capture the motion of the soliton with merely 17 mesh points (that is, $N = 15$). The trajectories of the mesh points for several simulations are depicted in Figure 9 and Figure 10. An example solution computed on a uniform mesh is depicted in Figure 8.

For the convergence test, we performed simulations for several N in the range 15-127. For comparison, we also computed solutions on a uniform mesh for N in the range 15-361. The numerical solutions were compared against the solution (168). The L^∞ errors are depicted in Figure 11. The L^∞ norms were evaluated over all nodes and over all time steps. Note that in case of a uniform mesh the spacing between the nodes is $\Delta x = X_{max}/(N+1)$, therefore the errors are plotted versus $(N+1)$. The Lagrange multiplier strategy proved to be more accurate than the control-theoretic strategy. As the number of mesh points is increased, the uniform mesh solution becomes quadratically convergent, as expected, since we used linear finite elements for spatial discretization. The control-theoretic strategy also shows near quadratic convergence, whereas the Lagrange multiplier method seems to converge slightly slower. While there are very few analytical results regarding the convergence of r-adaptive methods, it has been observed that the rate of convergence depends on several factors, including the chosen mesh density function. Our results are consistent with the convergence rates reported in [45] and [46]. Both papers deal with the viscous Burgers' equation, but consider different initial conditions. Computations with the arclength density function converged only linearly in [45], but quadratically in [46].

5.4. Energy conservation

As we pointed out in Section 2.6, the true power of variational and symplectic integrators for mechanical systems lies in their excellent conservation of energy and other integrals of motion, even when a big time step is used. In order to test the energy behavior of our methods, we performed simulations of the Sine-Gordon equation over longer time intervals. We considered two solitons

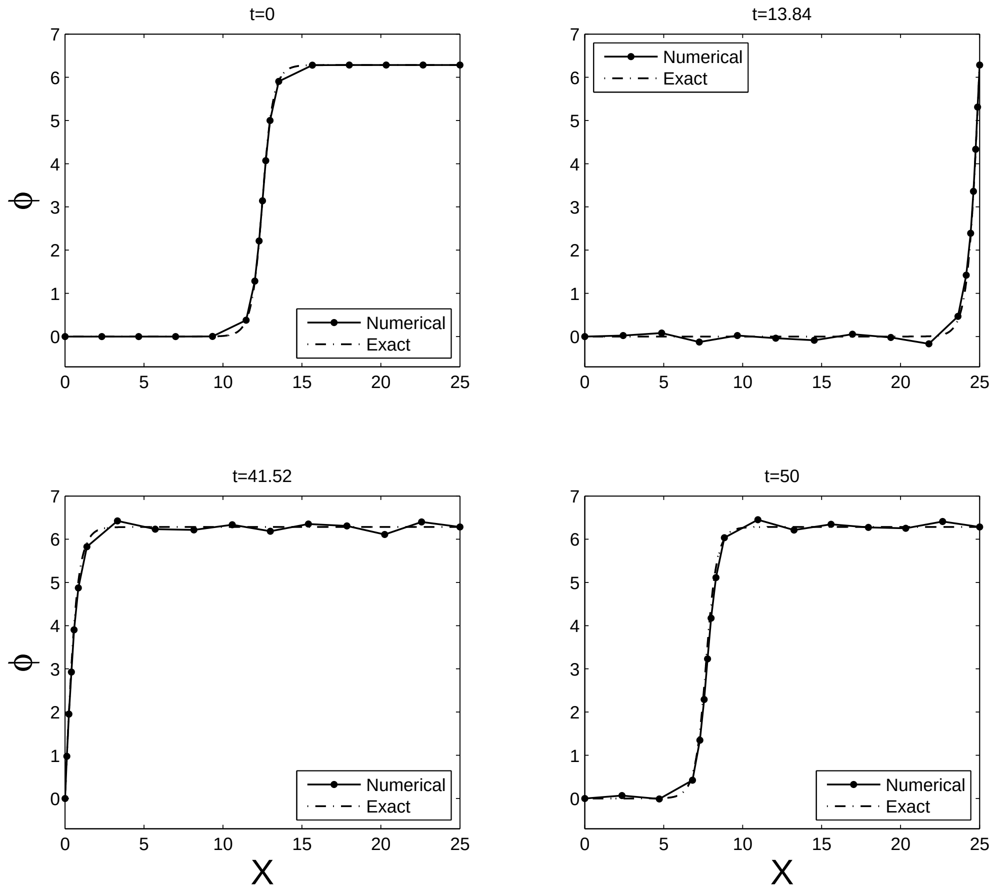

Figure 4. The single soliton solution obtained with the Lagrange multiplier strategy for $N = 15$. Integration in time was performed using the 4-th order Lobatto IIIA-IIIB scheme for constrained mechanical systems. The soliton moves to the right with the initial velocity $v = 0.9$, bounces from the right wall at $t = 13.84$ and starts moving to the left with the velocity $v = -0.9$, towards the left wall, from which it bounces at $t = 41.52$.

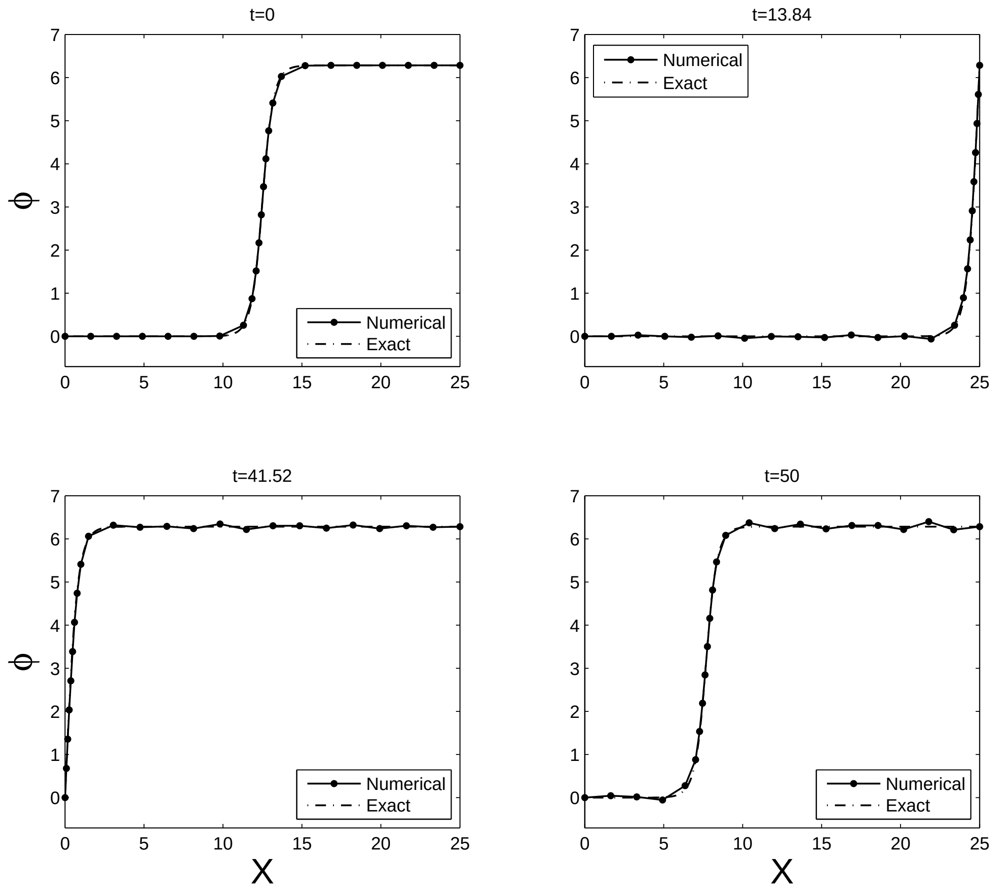

Figure 5. The single soliton solution obtained with the Lagrange multiplier strategy for $N = 31$. Integration in time was performed using the 4-th order Lobatto IIIA-IIIB scheme for constrained mechanical systems.

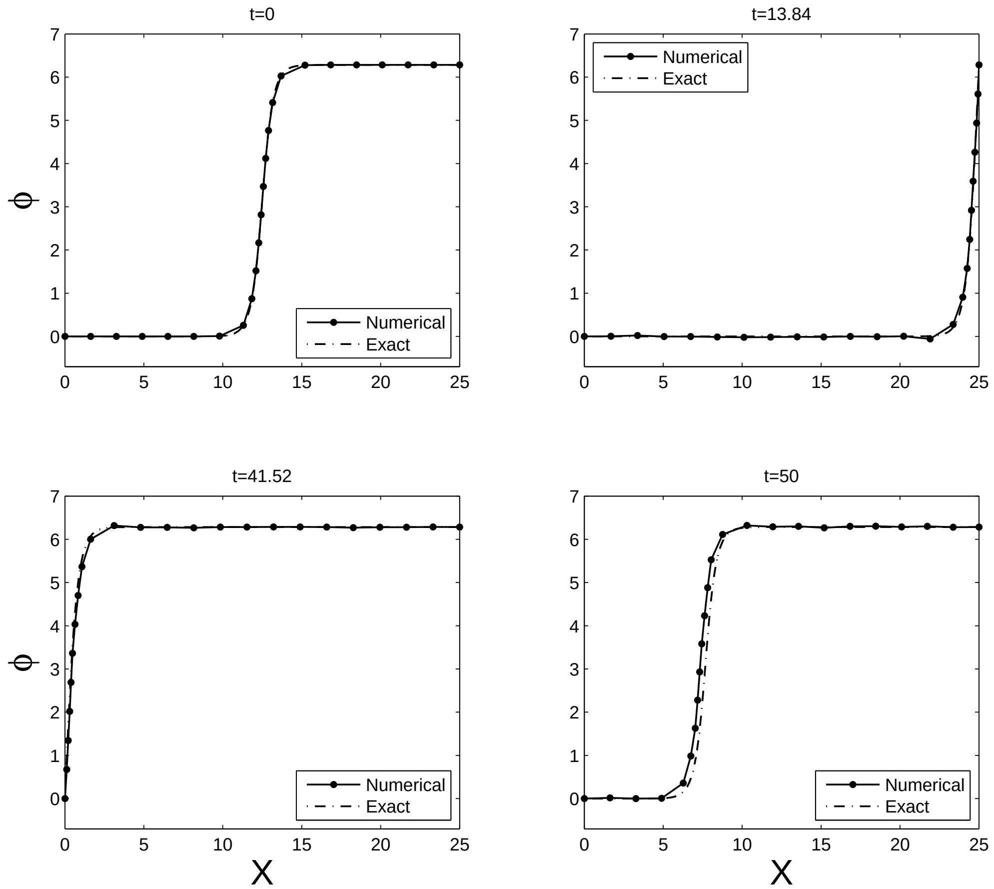

Figure 6. The single soliton solution obtained with the control-theoretic strategy for $N = 22$. Integration in time was performed using the 4-th order Gauss scheme. Integration with the 4-th order Lobatto IIIA-IIIB yields a very similar level of accuracy.

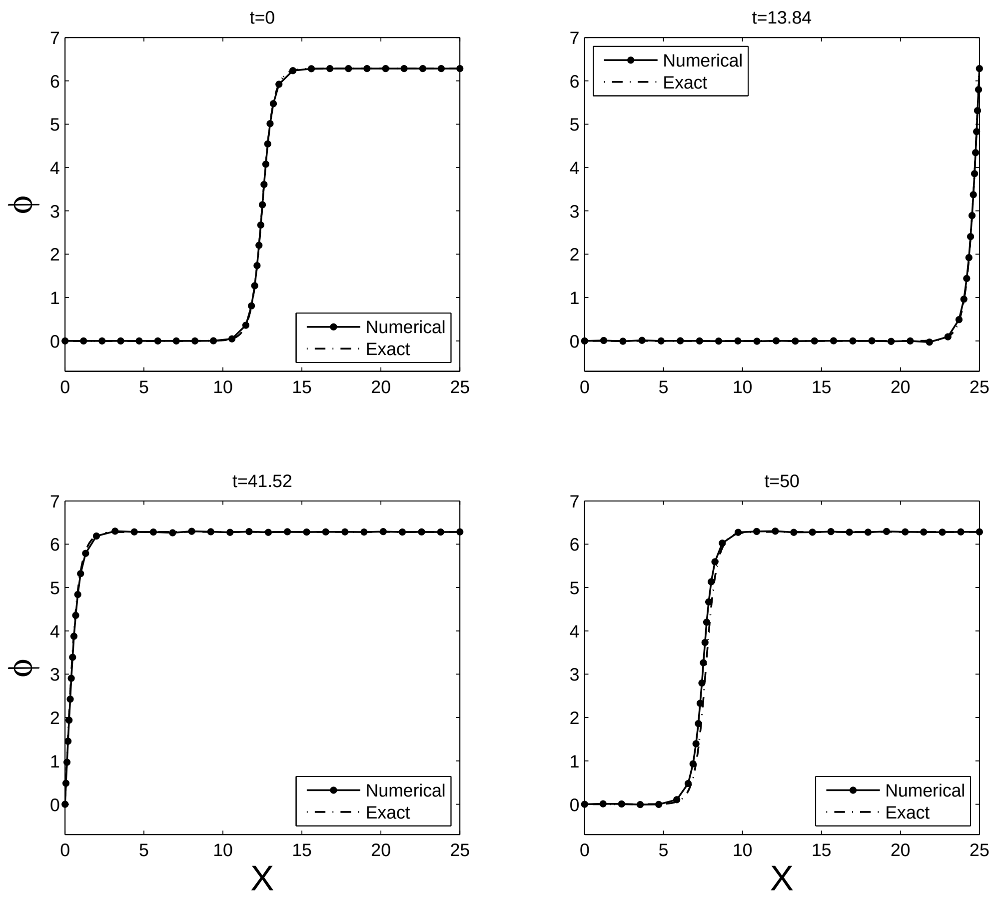

Figure 7. The single soliton solution obtained with the control-theoretic strategy for $N = 31$. Integration in time was performed using the 4-th order Gauss scheme. Integration with the 4-th order Lobatto IIIA-IIIB yields a very similar level of accuracy.

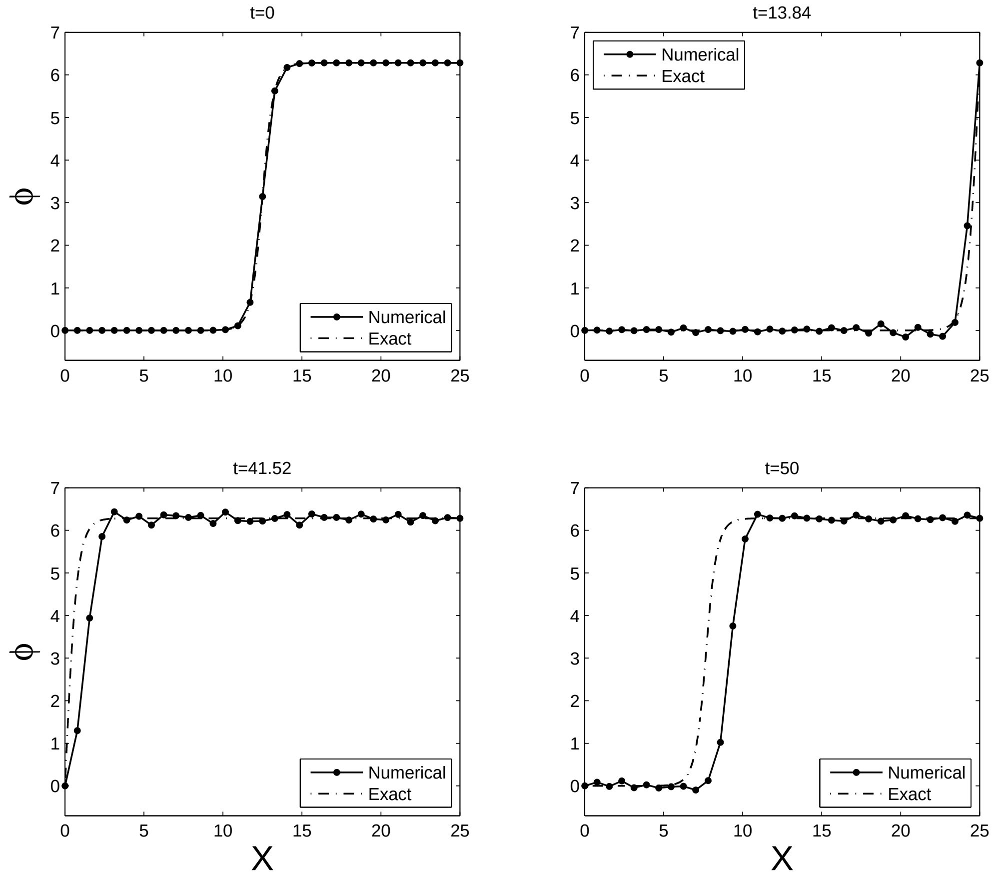

Figure 8. The single soliton solution computed on a uniform mesh with $N = 31$. Integration in time was performed using the 4-th order Gauss scheme. Integration with the 4-th order Lobatto IIIA-IIIB yields a very similar level of accuracy.

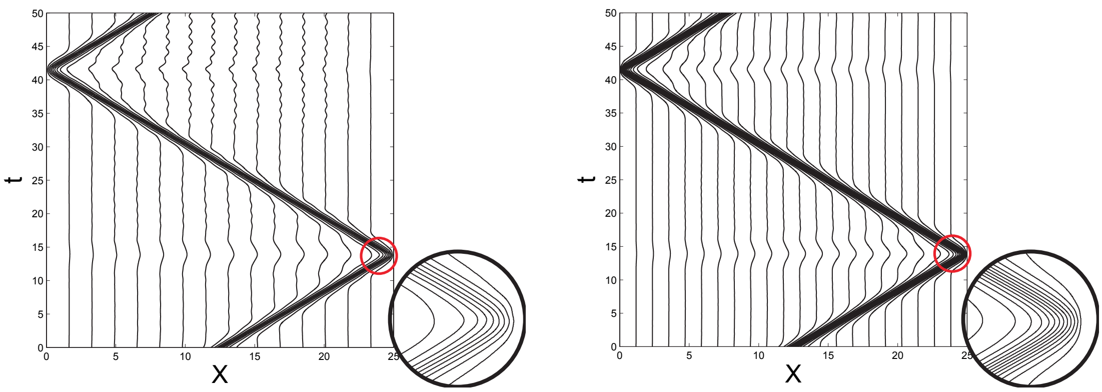

Figure 9. The mesh point trajectories (with zoomed-in insets) for the Lagrange multiplier strategy for $N = 22$ (left) and $N = 31$ (right). Integration in time was performed using the 4-th order Lobatto IIIA-IIIB scheme for constrained mechanical systems.

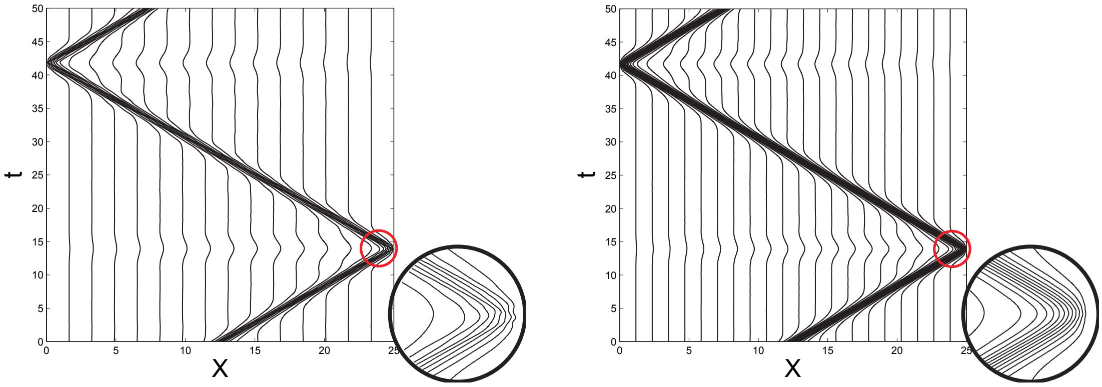

Figure 10. The mesh point trajectories (with zoomed-in insets) for the control-theoretic strategy for $N = 22$ (left) and $N = 31$ (right). Integration in time was performed using the 4-th order Gauss scheme. Integration with the 4-th order Lobatto IIIA-IIIB yields a very similar result.

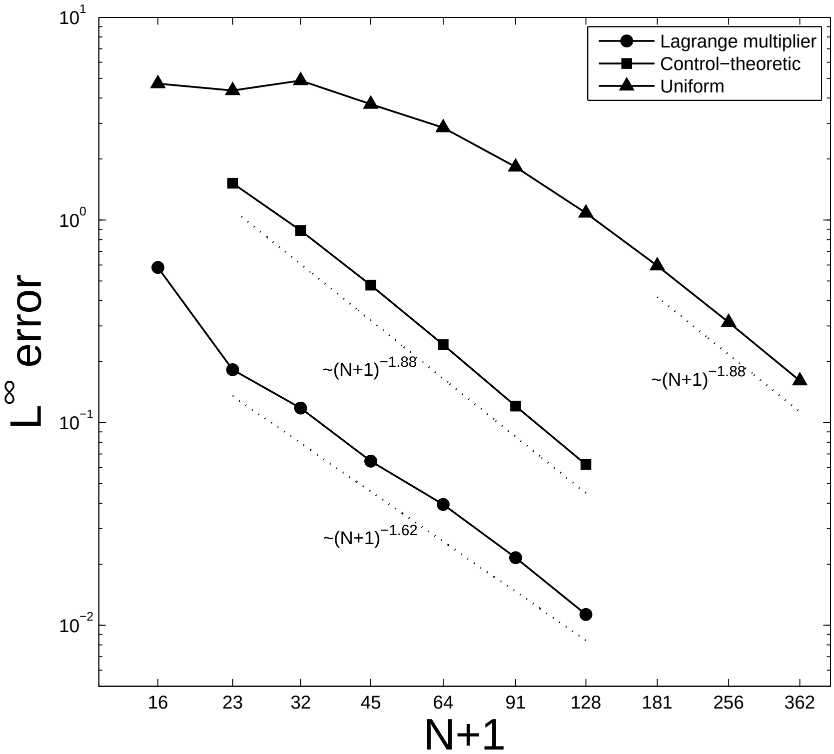

Figure 11. Comparison of the convergence rates of the discussed methods. Integration in time was performed using the 4-th order Lobatto IIIA-IIIB method for constrained systems in case of the Lagrange multiplier strategy, and the 4-th order Gauss scheme in case of both the control-theoretic strategy and the uniform mesh simulation. The 4-th order Lobatto IIIA-IIIB scheme for the control-theoretic strategy and the uniform mesh simulation yields a very similar level of accuracy. Also, using 2-nd order integrators gives very similar error plots.

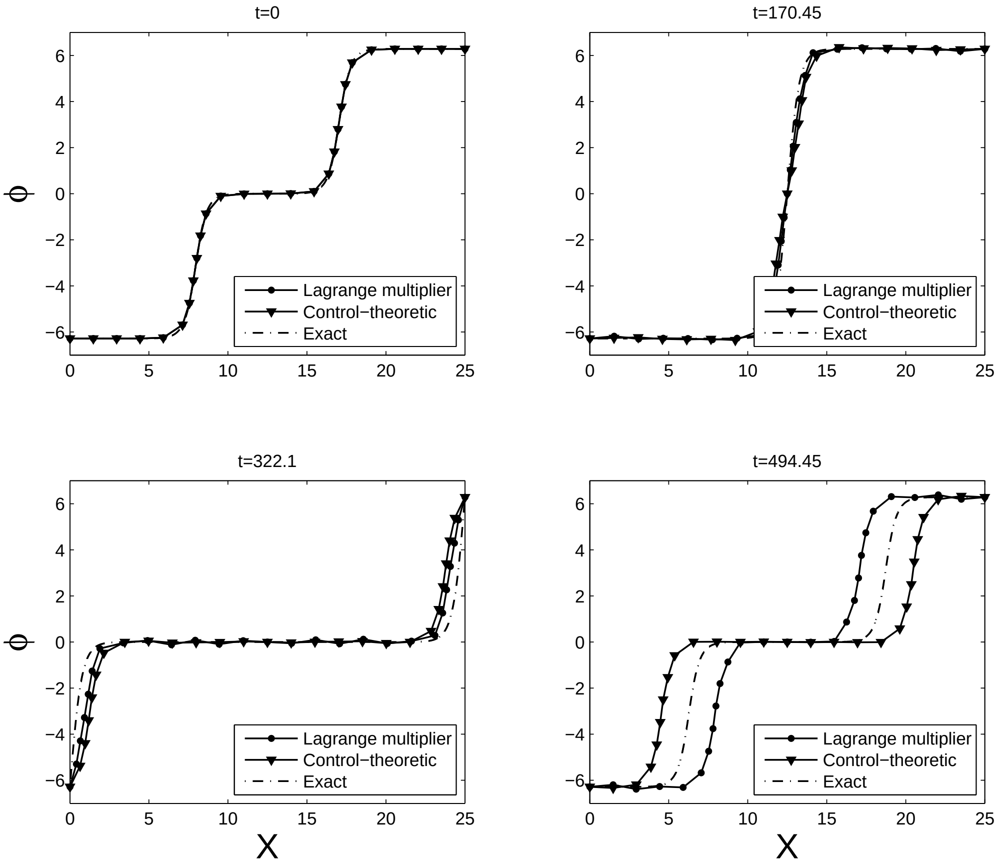

Figure 12. The two-soliton solution obtained with the control-theoretic and Lagrange multiplier strategies for $N = 25$. Integration in time was performed using the 4-th order Gauss quadrature for the control-theoretic approach, and the 4-th order Lobatto IIIA-IIIB quadrature for constrained mechanical systems in case of the Lagrange multiplier approach. The solitons initially move towards each other with the velocities $v = 0.9$, then bounce off of each other at $t = 5$ and start moving towards the walls, from which they bounce at $t = 18.79$. The solitons bounce off of each other again at $t = 32.57$. This solution is periodic in time with the period $T_{period} = 27.57$. The nearly exact solution was constructed in a similar fashion as (168). As the simulation progresses, the Lagrange multiplier solution gets ahead of the exact solution, whereas the control-theoretic solution lags behind.

bouncing from each other and from two rigid walls at $X = 0$ and $X_{max} = 25$. We imposed the boundary conditions $\phi_L = -2\pi$ and $\phi_R = 2\pi$, and as initial conditions we used $\phi(X, 0) = \phi_{SS}(X - 12.5, -5)$ with $v = 0.9$. We ran our computations on a mesh consisting of 27 nodes (N=25). Integration was performed with the time step $\Delta t = 0.05$, which is rather large for this type of simulations. The scaling parameter in (28) was set to $\alpha = 1.5$, so that approximately half of the available mesh points were concentrated in the areas of high gradient. An example solution is presented in Figure 12.

The exact energy of the two-soliton solution can be computed using (7). It is possible to compute that integral explicitly to obtain $E = 16/\sqrt{1 - v^2} \approx 36.71$. The energy associated with the semi-discrete Lagrangian (44) can be expressed by the formula

$$E_N = \frac{1}{2}\dot{q}^T \tilde{M}_N(q)\,\dot{q} + R_N(q), \tag{169}$$

where R_N was defined in (88) and for our Sine-Gordon system is given by

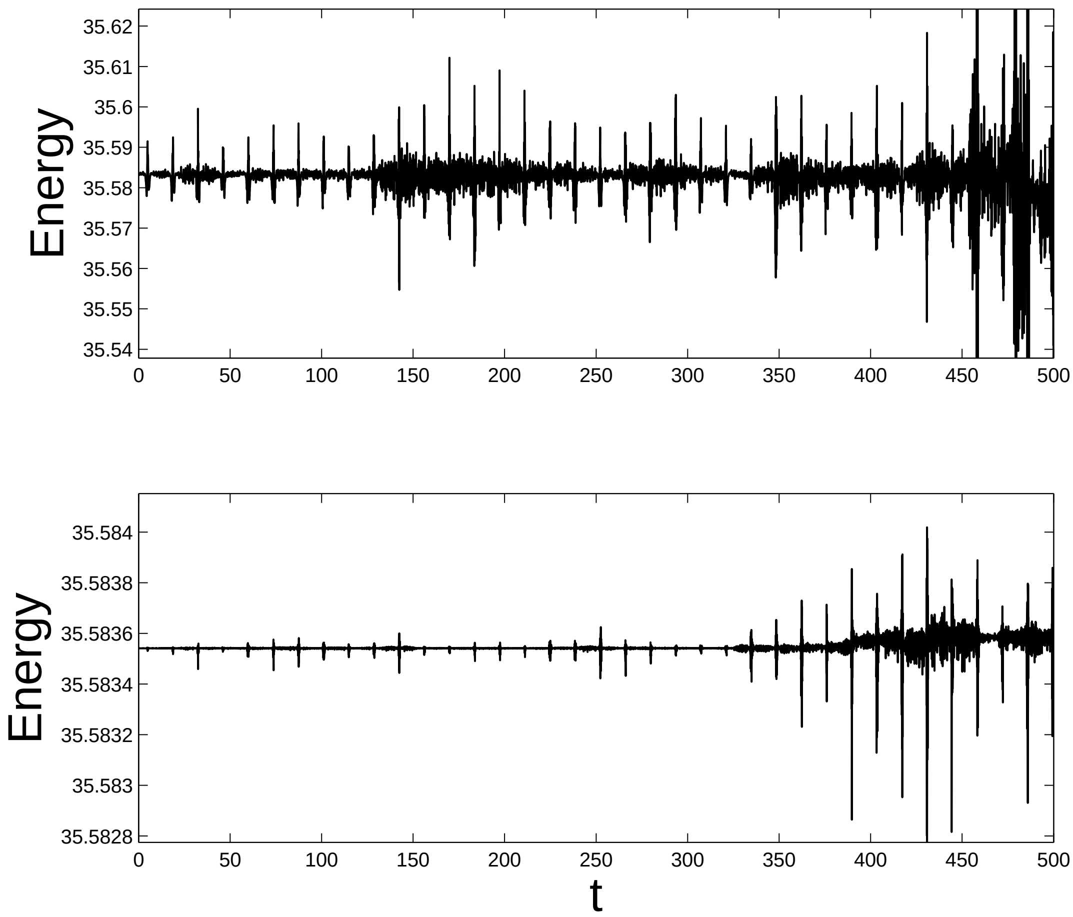

Figure 13. The discrete energy E_N for the Lagrange multiplier strategy. Integration in time was performed with the 2-nd (top) and 4-th (bottom) order Lobatto IIIA-IIIB method for constrained mechanical systems. The spikes correspond to the times when the solitons bounce off of each other or of the walls.

$$R_N(q) = \sum_{k=0}^{N} \left[\frac{1}{2} \left(\frac{y_{k+1} - y_k}{X_{k+1} - X_k} \right)^2 + 1 - \frac{\sin y_{k+1} - \sin y_k}{y_{k+1} - y_k} \right] (X_{k+1} - X_k), \tag{170}$$

and M_N is the mass matrix (47). The energy E_N is an approximation to (7) if the integrand is sampled at the nodes $X_0, \ldots, X_{N+1}$ and then piecewise linearly approximated. Therefore, we used E_N to compute the energy of our numerical solutions.

The energy plots for the Lagrange multiplier strategy are depicted in Figure 13. We can see that the energy stays nearly constant in the presented time interval, showing only mild oscillations, which are reduced as higher order of integration in time is used. The energy plots for the control-theoretic strategy are depicted in Figure 14. In this case the discrete energy is more erratic and not as nearly preserved. Moreover, the symplectic Gauss and Lobatto methods show virtually the same energy behavior as the non-symplectic Radau IIA method, which is known for its excellent stability properties when applied to stiff differential equations (see [12]). It seems that we do not gain much by performing symplectic integration in this case. It is consistent with our observations in Section 2.6 and shows that the control-theoretic strategy does not take the full advantage of the underlying geometry.

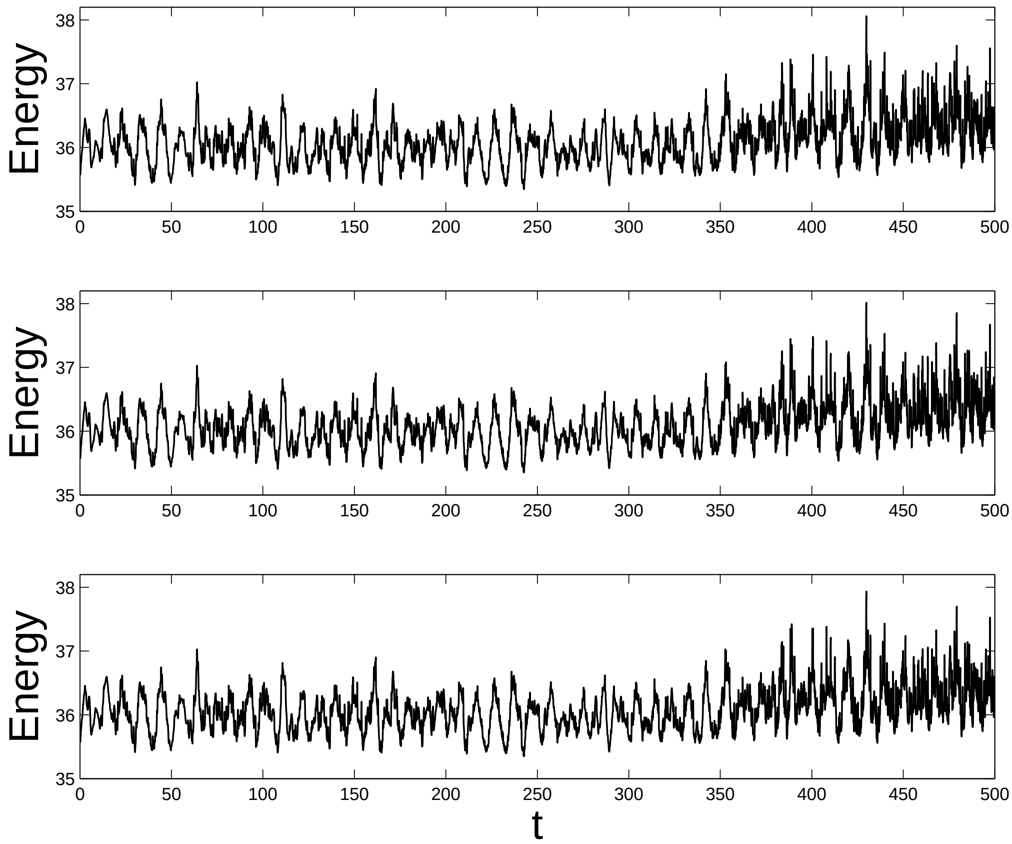

Figure 14. The discrete energy E_N for the control-theoretic strategy. Integration in time was performed with the 4-th order Gauss (top), 4-th order Lobatto IIIA-IIIB (middle) and non-symplectic 5-th order Radau IIA (bottom) methods.

As we did not use adaptive time-stepping and did not implement any mesh smoothing techniques, the quality of the mesh deteriorated with time in all the simulations, eventually leading to mesh crossing, i.e. two mesh points collapsing or crossing each other. The control-theoretic strategy, even though less accurate, retained good mesh quality longer, with the break-down time $T_{break} > 1000$, as opposed to $T_{break} \sim 600$ in case of the Lagrange multiplier approach (both using a rather large constant time step). We discuss extensions to our approach for increased robustness in Section 6.

6. Summary and future work

We have proposed two general ideas how *r*-adaptive meshes can be applied in geometric numerical integration of Lagrangian partial differential equations. We have constructed several variational and multisymplectic integrators and discussed their properties. We have used the Sine-Gordon model and its solitonic solutions to test our integrators numerically.

Our work can be extended in many directions. Interestingly, it also opens many questions in geometric mechanics and multisymplectic field theory. Addressing those questions may have a broad impact on the field of geometric numerical integration.

Non-hyperbolic equations

The special form of the Lagrangian density (42) we considered leads to a hyperbolic PDE, which poses a challenge to *r*-adaptive methods, as at each time step the mesh is adapted *globally* in response to *local* changes in the solution. *Causality* and the structure of the characteristic lines of hyperbolic systems make *r*-adaptation prone to instabilities and integration in time has to be performed carefully. The literature on *r*-adaptation almost entirely focuses on parabolic problems (see [7], [8] and references therein). Therefore, it would be interesting to apply our methods to PDEs that are first-order in time, for instance the Korteweg-de Vries, Nonlinear Schrödinger or Camassa-Holm equations. All three equations are first-order in time and are not hyperbolic in nature. Moreover, all can be derived as Lagrangian field theories (see [47], [48], [49], [42], [50], [51], [35]). The Nonlinear Schrödinger equation has applications to optics and water waves, whereas the Korteweg-de Vries and Camassa-Holm equations were introduced as models for waves in shallow water. All equations possess interesting solitonic solutions. The purpose of *r*-adaptation would be to improve resolution, for instance, to track the motion of solitons by placing more mesh points near their centers and making the mesh less dense in the asymptotically flat areas.

Hamiltonian Field Theories

Variational multisymplectic integrators for field theories have been developed in the Lagrangian setting ([35], [3]). However, many interesting field theories are formulated in the Hamiltonian setting. They may not even possess a Lagrangian formulation. It would be interesting to construct Hamiltonian variational integrators for multisymplectic PDEs by generalizing the variational characterization of discrete Hamiltonian mechanics. This would allow to handle Hamiltonian PDEs without the need for converting them to the Lagrangian framework. Recently Leok & Zhang [52] and Vankerschaver & Ciao & Leok [53] have laid foundations for such integrators. It would also be interesting to see if the techniques we used in our work could be applied in order to construct *r*-adaptive Hamiltonian integrators.

Time adaptation based on local error estimates

One of the challenges of *r*-adaptation is that it requires solving differential-algebraic or stiff ordinary differential equations. This is because there are two different time scales present: one defined by the physics of the problem and one following from the strategy we use to adapt the mesh. Stiff ODEs and DAEs are known to require time integration with an adaptive step size control based on local error estimates (see [11], [12]). In our work we used constant time-stepping, as adaptive step size control is difficult to combine with geometric numerical integration. Classical step size control is based on past information only, time symmetry is destroyed and with it the qualitative properties of the method. Hairer & Söderlind [54] developed explicit, reversible, symmetry-preserving, adaptive step size selection algorithms for geometric integrators, but their method is not based on local error estimation, thus it is not useful for r-adaptation. Symmetric error estimators are considered in [28] and some promising results are discussed. Hopefully, the ideas presented in those papers could be combined and generalized. The idea of Asynchronous Variational Integrators (see [4]) could also be useful here, as this would allow to use a different time step for each cell of the mesh.

Constrained multisymplectic field theories

The multisymplectic form formula (106) was first introduced in [3]. The authors, however, consider only unconstrained field theories. In our work we start with the unconstrained field theory (1), but upon choosing an adaptation strategy represented by the constraint $G = 0$ we obtain a constrained theory, as described in Section 3 and Section 4.3. Moreover, this constraint is essentially nonholonomic, as it contains derivatives of the fields, and the equations of motion are obtained using the *vakonomic* approach (also called variational nonholonomic) rather than the Lagrange-d'Alembert principle. All that gives rise to many very interesting and general questions. Is there a multisymplectic form formula for such theories? Is it derived in a similar fashion? Do variational integrators obtained this way satisfy some discrete multisymplectic form formula? These issues have been touched upon in [41], but by no means resolved.

Mesh smoothing and variational nonholonomic integrators

The major challenge of r-adaptive methods is *mesh crossing*, which occurs when two mesh points collapse or cross each other. In order to avoid mesh crossing and retain good mesh quality, mesh smoothing techniques were developed ([7], [8]). They essentially attempt to regularize the exact equidistribution constraint $G = 0$ by replacing it with the condition $\epsilon\, \partial X/\partial t = G$, where ϵ is a small parameter. This can be interpreted as adding some attraction and repulsion pseudoforces between mesh points. If one applies the Lagrange multiplier approach to r-adaptation as described in Section 3, then upon finite element discretization one obtains a finite dimensional Lagrangian system with a nonholonomic constraint. This constraint is enforced using the vakonomic (nonholonomic variational) formulation. Variational integrators for systems with nonholonomic constraints have been developed mostly in the Lagrange-d'Alembert setting, but there have also been some results regarding discrete vakonomic mechanics. The ideas presented in [55], [56], and [57] may be used to design structure-preserving mesh smoothing techniques.

Author Contributions: Our contributions were equally balanced in a true collaboration.

Funding: This research received no external funding.

Acknowledgments: We would like to extend our gratitude to Michael Holst, Eva Kanso, Patrick Mullen, Tudor Ratiu, Ari Stern and Abigail Wacher for useful comments and suggestions. We are particularly indebted to Joris Vankerschaver and Melvin Leok for support, discussions and interest in this work. We dedicate this paper in memory of Jerrold E. Marsden, who began this project with us.

Conflicts of Interest: The authors declare no conflict of interest.

1. Hairer, E.; Lubich, C.; Wanner, G. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*; Springer Series in Computational Mathematics, Springer, New York, 2002.
2. Marsden, J.E.; West, M. Discrete mechanics and variational integrators. *Acta Numerica* **2001**, *10*, 357–514.
3. Marsden, J.E.; Patrick, G.W.; Shkoller, S. Multisymplectic geometry, variational integrators, and nonlinear PDEs. *Communications in Mathematical Physics* **1998**, *199*, 351–395.
4. Lew, A.; Marsden, J.E.; Ortiz, M.; West, M. Asynchronous variational integrators. *Archive for Rational Mechanics and Analysis* **2003**, *167*, 85–146.

5. Stern, A.; Tong, Y.; Desbrun, M.; Marsden, J.E. Variational integrators for Maxwell's equations with sources. *PIERS Online* **2008**, *4*, 711–715. doi:10.2529/PIERS071019000855.
6. Pavlov, D.; Mullen, P.; Tong, Y.; Kanso, E.; Marsden, J.E.; Desbrun, M. Structure-preserving discretization of incompressible fluids. *Physica D: Nonlinear Phenomena* **2011**, *240*, 443–458.
7. Budd, C.J.; Huang, W.; Russell, R.D. Adaptivity with moving grids. *Acta Numerica* **2009**, *18*, 111–241. doi:10.1017/S0962492906400015.
8. Huang, W.; Russell, R. *Adaptive Moving Mesh Methods*; Vol. 174, *Applied Mathematical Sciences*, Springer Verlag, 2011.
9. Nijmeijer, H.; van der Schaft, A. *Nonlinear Dynamical Control Systems*; Springer, New York, 1990.
10. Gotay, M. Presymplectic manifolds, geometric constraint theory and the Dirac-Bergmann theory of constraints. PhD thesis, University of Maryland, College Park, 1979.
11. Brenan, K.; Campbell, S.; Petzold, L. *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*; Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, 1996.
12. Hairer, E.; Wanner, G. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*, 2nd ed.; Vol. 14, *Springer Series in Computational Mathematics*, Springer, 1996.
13. Hairer, E.; Lubich, C.; Roche, M. *The numerical solution of differential-algebraic systems by Runge-Kutta methods*; Lecture Notes in Math. 1409, Springer Verlag, 1989.
14. Hairer, E.; Nørsett, S.; Wanner, G. *Solving Ordinary Differential Equations I: Nonstiff Problems*, 2nd ed.; Vol. 8, *Springer Series in Computational Mathematics*, Springer, 1993.
15. Ebin, D.G.; Marsden, J. Groups of Diffeomorphisms and the Motion of an Incompressible Fluid. *Annals of Mathematics* **1970**, *92*, 102–163.
16. Evans, L. *Partial Differential Equations*; Graduate studies in mathematics, American Mathematical Society, 2010.
17. Rabier, P.J.; Rheinboldt, W.C. Theoretical and Numerical Analysis of Differential-Algebraic Equations. In *Handbook of Numerical Analysis*; Ciarlet, P.G.; Lion, J.L., Eds.; Elsevier Science B.V., 2002; Vol. 8, pp. 183–540.
18. Reißig, G.; Boche, H. On singularities of autonomous implicit ordinary differential equations. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on* **2003**, *50*, 922–931.
19. Rabier, P.J. Implicit differential equations near a singular point. *Journal of Mathematical Analysis and Applications* **1989**, *144*, 425–449.
20. Rabier, P.J.; Rheinboldt, W.C. A general existence and uniqueness theory for implicit differential-algebraic equations. *J. Diff. and Integral Equations* **1991**, *4*, 563–582.
21. Rabier, P.J.; Rheinboldt, W.C. A geometric treatment of implicit differential-algebraic equations. *Journal of Differential Equations* **1994**, *109*, 110–146.
22. Rabier, P.J.; Rheinboldt, W.C. On impasse points of quasilinear differential-algebraic equations. *J. Math. Anal. Appl.* **1994**, *181*, 429–454.
23. Rabier, P.J.; Rheinboldt, W.C. On the computation of impasse points of quasilinear differential-algebraic equations. *Mathematics of Computation* **1994**, *62*, 133–154.
24. Miller, K.; Miller, R.N. Moving finite elements I. *SIAM Journal on Numerical Analysis* **1981**, *18*, 1019–1032.
25. Miller, K. Moving finite elements II. *SIAM Journal on Numerical Analysis* **1981**, *18*, 1033–1057.
26. Zielonka, M.; Ortiz, M.; Marsden, J. Variational r-adaption in elastodynamics. *International Journal for Numerical Methods in Engineering* **2008**, *74*, 1162–1197.
27. Jay, L. Symplectic partitioned Runge-Kutta methods for constrained Hamiltonian systems. *SIAM Journal on Numerical Analysis* **1996**, *33*, 368–387.
28. Jay, L.O. Structure Preservation for Constrained Dynamics with Super Partitioned Additive Runge–Kutta Methods. *SIAM Journal on Scientific Computing* **1998**, *20*, 416–446.
29. Leimkuhler, B.J.; Skeel, R.D. Symplectic Numerical Integrators in Constrained Hamiltonian Systems. *Journal of Computational Physics* **1994**, *112*, 117 – 125. doi:https://doi.org/10.1006/jcph.1994.1085.
30. Leimkuhler, B.; Reich, S. *Simulating Hamiltonian Dynamics*; Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 2004.
31. Leyendecker, S.; Marsden, J.; Ortiz, M. Variational integrators for constrained dynamical systems. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik* **2008**, *88*, 677–708. doi:10.1002/zamm.200700173.

32. Marsden, J.; Ratiu, T. *Introduction to Mechanics and Symmetry*; Vol. 17, *Texts in Applied Mathematics*, Springer Verlag, 1994.
33. Saunders, D. *The Geometry of Jet Bundles*; Vol. 142, *London Mathematical Society Lecture Note Series*, Cambridge University Press, 1989.
34. Gotay, M.; Isenberg, J.; Marsden, J.; Montgomery, R. Momentum Maps and Classical Relativistic Fields. Part I: Covariant Field Theory. Unpublished, arXiv:physics/9801019.
35. Kouranbaeva, S.; Shkoller, S. A variational approach to second-order multisymplectic field theory. *Journal of Geometry and Physics* **2000**, *35*, 333–366.
36. Gotay, M. A multisymplectic framework for classical field theory and the calculus of variations I: covariant Hamiltonian formulation. In *Mechanics, analysis and geometry: 200 years after Lagrange*; Francavigila, M., Ed.; North-Holland, Amsterdam, 1991; pp. 203–235.
37. Bloch, A. *Nonholonomic Mechanics and Control*; Interdisciplinary Applied Mathematics, Springer, 2003.
38. Bloch, A.M.; Crouch, P.E. Optimal Control, Optimization, and Analytical Mechanics. In *Mathematical Control Theory*; Baillieul, J.; Willems, J., Eds.; Springer New York, 1999; pp. 268–321. doi:10.1007/978-1-4612-1416-8_8.
39. Bloch, A.M.; Krishnaprasad, P.; Marsden, J.E.; Murray, R.M. Nonholonomic mechanical systems with symmetry. *Archive for Rational Mechanics and Analysis* **1996**, *136*, 21–99. doi:10.1007/BF02199365.
40. Cortés, J.; de León, M.; de Diego, D.; Martínez, S. Geometric Description of Vakonomic and Nonholonomic Dynamics. Comparison of Solutions. *SIAM Journal on Control and Optimization* **2002**, *41*, 1389–1412. doi:10.1137/S036301290036817X.
41. Marsden, J.E.; Pekarsky, S.; Shkoller, S.; West, M. Variational methods, multisymplectic geometry and continuum mechanics. *Journal of Geometry and Physics* **2001**, *38*, 253–284.
42. Drazin, P.; Johnson, R. *Solitons: An Introduction*; Cambridge Computer Science Texts, Cambridge University Press, 1989.
43. Rajaraman, R. *Solitons and Instantons: An Introduction to Solitons and Instantons in Quantum Field Theory*; North-Holland personal library, North-Holland Publishing Company: Amsterdam, 1982.
44. Allgower, E.; Georg, K. *Introduction to Numerical Continuation Methods*; Classics in applied mathematics, Society for Industrial and Applied Mathematics, 2003.
45. Beckett, G.; Mackenzie, J.; Ramage, A.; Sloan, D. On The Numerical Solution of One-Dimensional PDEs Using Adaptive Methods Based on Equidistribution. *Journal of Computational Physics* **2001**, *167*, 372 – 392. doi:http://dx.doi.org/10.1006/jcph.2000.6679.
46. Wacher, A. A comparison of the String Gradient Weighted Moving Finite Element method and a Parabolic Moving Mesh Partial Differential Equation method for solutions of partial differential equations. *Central European Journal of Mathematics* **2013**, *11*, 642–663.
47. Camassa, R.; Holm, D.D. An integrable shallow water equation with peaked solitons. *Phys. Rev. Lett.* **1993**, *71*, 1661–1664. doi:10.1103/PhysRevLett.71.1661.
48. Camassa, R.; Holm, D.D.; Hyman, J. A new integrable shallow water equation. *Adv. App. Mech.* **1994**, *31*, 1–31.
49. Chen, J.B.; Qin, M.Z. A multisymplectic variational integrator for the nonlinear Schrödinger equation. *Numerical Methods for Partial Differential Equations* **2002**, *18*, 523–536. doi:10.1002/num.10021.
50. Faou, E. *Geometric Numerical Integration and Schrödinger Equations*; Zurich lectures in advanced mathematics, European Mathematical Society, 2012.
51. Gotay, M. A multisymplectic approach to the KdV equation. Differential Geometric Methods in Theoretical Physics. NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences, 1988, Vol. 250, pp. 295–305.
52. Leok, M.; Zhang, J. Discrete Hamiltonian variational integrators. *IMA Journal of Numerical Analysis* **2011**, *31*, 1497–1532.
53. Vankerschaver, J.; Leok, M. A novel formulation of point vortex dynamics on the sphere: geometrical and numerical aspects. *J. Nonlin. Sci.* **2014**, *24*, 1–37.
54. Hairer, E.; Söderlind, G. Explicit, time reversible, adaptive step size control. *SIAM J. Sci. Comput.* **2005**, *26*, 1838–1851.
55. Benito, R.; Martín de Diego, D. Discrete vakonomic mechanics. *Journal of Mathematical Physics* **2005**, *46*, 083521. doi:10.1063/1.2008214.

56. García, P.L.; Fernández, A.; Rodrigo, C. Variational integrators in discrete vakonomic mechanics. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas* **2012**, *106*, 137–159. doi:10.1007/s13398-011-0030-x.
57. Colombo, L.; Martín de Diego, D.; Zuccalli, M. Higher-order discrete variational problems with constraints. *Journal of Mathematical Physics* **2013**, *54*, 093507. doi:10.1063/1.4820817.