High-Order Symplectic Partitioned Lie Group Methods

Geir Bogfjellmo · Håkon Marthinsen

Abstract In this article, a unified approach to obtain symplectic integrators on $T^*G$ from Lie group integrators on a Lie group $G$ is presented. The approach is worked out in detail for symplectic integrators based on Runge–Kutta–Munthe-Kaas methods and Crouch–Grossman methods. These methods can be interpreted as symplectic partitioned Runge–Kutta methods extended to the Lie group setting in two different ways. In both cases, we show that it is possible to obtain symplectic integrators of arbitrarily high order by this approach.

Keywords Symplectic integrators · Lie groups · Order theory

Mathematics Subject Classification Primary 65P10; Secondary 37M15 · 70G65 · 70G75 · 70HXX

1 Introduction

1.1 Motivation and Background

In general, an ordinary differential equation (ODE) can be described by a vector field on a smooth manifold where solutions of the ODE are integral curves of the vector
Numerical approximation of solutions of ODEs is an old field of study, and a plethora of methods for obtaining numerical solutions exists. However, most of these methods assume that the manifold is Euclidean space. If the manifold is not Euclidean space, it is possible to embed the manifold in Euclidean space and extend the vector field on the manifold to a vector field in Euclidean space such that the integral curves are ensured to remain in the image of the embedding. A standard numerical algorithm (e.g., a Runge–Kutta method) will in general result in discrete points which do not lie in the image of the embedding. An improvement of this approach is to use projection methods to obtain solutions on the manifold. These approaches, though simple, suffer from the problem that the numerical solutions depend on the particular choice of embedding and on the particular extension of the vector field.

One aspect of geometric numerical integration is to exploit structure on the manifold to define numerical methods that are intrinsic to the manifold (i.e., do not depend on a particular embedding). This structure can for instance be that of a Lie group acting on the manifold. The action of a Lie group $G$ on a manifold $M$ is a smooth mapping $\Psi: G \times M \to M$ which respects the group structure on $G$. If the action is transitive, then the derivative with respect to the first component of $\Psi$ at the group identity $e$ is a surjective vector bundle morphism $g \times M \to TM$. Any vector field $X$ on $M$ can then be lifted (possibly in a non-unique manner) to a section of the vector bundle $g \times M$. The combination of this lifting and standard charts $g \to G$ forms the basis of several classes of Lie group methods. Among them are the Crouch–Grossman (CG) methods [1] and the Runge–Kutta–Munthe-Kaas (RKMK) methods [2]. For a more detailed discussion of Lie group methods, we refer to the survey article by Iserles et al. [3] and the references therein.

Another aspect of geometric numerical integration is symplecticity of numerical integrators. Many important problems from physics can be formulated as Hamiltonian ODEs on cotangent bundles over manifolds. The flow maps of these ODEs are symplectic, that is, they preserve the canonical two-form on the cotangent bundle. For Hamiltonian ODEs, it is beneficial to use symplectic integrators, due to the near-preservation of energy and excellent long-term behavior of the numerical solutions [4, Chapter VI]. Hamilton’s principle states that the solution of a Lagrangian (in many cases also Hamiltonian) system moves along a path which extremizes the action integral $S = \int_0^T L(q(t), \dot{q}(t)) \, dt$ among all paths $q$ with fixed end points.

One technique for deriving symplectic methods is based on the notion of discretizing Hamilton’s principle, that is, replacing the action integral with a discrete action sum and extremizing over all discrete paths or sequences of points $q_0, q_1, \ldots, q_N$ with fixed end points. These methods are known as variational methods or variational integrators. Variational methods are guaranteed to be symplectic since the terms $L_h(q_{k-1}, q_k)$ of the discrete action sum can be interpreted as generating functions (of the first type) for the numerical flow map. Variational methods have been studied by numerous authors, we refer to the review article by Marsden and West [5] or the more recent encyclopedia article by Leok [6] and the references therein for more information about variational methods.

Standard Lie group methods, such as RKMK methods or CG methods, give numerical solutions that evolve on the same manifolds as the exact solutions. The question
of the existence of symplectic methods of formats similar to the ones considered by Crouch and Grossman or by Munthe-Kaas has been a topic of interest for several years.

On \( \mathbb{R}^n \), there is a unified way to extend Runge–Kutta (RK) methods on \( \mathbb{R}^n \) to symplectic methods on \( T^*\mathbb{R}^n \) [4, Section VI.6.3], i.e., symplectic partitioned RK (SPRK) methods. Our goal with this article was to construct and study symplectic methods of arbitrarily high order that are extended from Lie group methods, i.e., high-order symplectic Lie group integrators. We have focused on the case where \( M = G \), and the action is simply multiplication in the Lie group. In the case where \( M \neq G \), isotropy complicates matters. The technique of extremizing a discrete action sum still yields symplectic mappings in the case \( M \neq G \), but the presence of isotropy complicates the analysis of these integrators. The details of the isotropy case will hopefully be addressed in a later article.

The idea of constructing variational methods from Lie group methods has previously been considered by several authors. Bou-Rabee and Marsden proposed in 2009 to base variational methods on RKMK methods [7] and present a class of methods of first and second orders. Methods of a similar type are also considered in the survey article by Celledoni, Marthinsen, and Owren [8]. In the present article, this idea is pursued further to obtain methods of arbitrarily high order.

It is known to the authors that a different, but related approach to variational Lie group methods has been studied by Leok and collaborators. Their approach is based on approximating the curve \( q \) in a finite-dimensional function space, resulting in Galerkin Lie group variational integrators. The idea appears already in Leok’s doctoral thesis [9, Section 5.3] and also in other articles by Leok and co-authors. A more detailed study of this approach can be found in an article by Hall and Leok [10].

The rest of Sect. 1 is an introduction to ODEs on a Lie group \( G \), the Hamilton–Pontryagin (HP) principle and the equivalent HP equations, and variational integrators in general. Section 2 begins by introducing a group structure on \( T^*G \), or equivalently, on \( G \times \mathfrak{g}^* \) and a function \( f : G \times \mathfrak{g}^* \to \mathfrak{g} \times \mathfrak{g}^* \) which together fully describe ODEs on \( T^*G \). Next, we introduce the general format for our integrators. In Sect. 3, we first show that a subclass of our integrators who have been studied before [7, 8] cannot obtain higher than second order on general Hamiltonian problems. We then show that our integrators cannot obtain higher order than the underlying Lie group integrators. In Sect. 4, we present two classes of higher-order integrators which are based on RKMK integrators and CG methods, respectively. In Sect. 5, we show that both classes of methods from Sect. 4 can obtain arbitrarily high order, and we present general order conditions for the methods based on RKMK integrators and conditions for orders 1–3 for the CG-based integrators. We test the two classes of methods numerically in Sect. 6 and show that they both achieve the correct order and that they both have small energy errors over long time. Finally, in Sect. 7, we conclude and mention some possible topics for further work.

1.2 ODEs on Lie Groups

Let \( G \) be a finite-dimensional Lie group and \( \mathfrak{g} \) its associated Lie algebra. We denote right multiplication with \( g \in G \) as \( R_g \) and left multiplication with \( g \) as \( L_g \).
dot notation to denote translation in the tangent bundle, i.e.,
\[ g \cdot v = TL_g v, \quad v \cdot g = TR_g v, \quad v \in TG, \]
and in the cotangent bundle
\[ g \cdot p = T^* L_{g^{-1}} p, \quad p \cdot g = T^* R_{g^{-1}} p, \quad p \in T^* G. \]
We also need the notation Ad\_g := TL\_g \circ TR\_g \_1. All autonomous ODEs on G can be written as
\[ \dot{g} = f(g) \cdot g, \quad g(0) = g_0, \quad (1) \]
where g is a curve in G, and the map \( f: G \to \mathfrak{g} \) is determined uniquely by the vector field. We can solve this kind of equation numerically using Lie group methods [3]. Here, we have chosen the right-trivialized form of this equation. We could also have used the left-trivialized form \( \dot{g} = g \cdot f(g) \) which would have resulted in only minor changes to the formulae presented later in the article.

Since we are interested in solving Hamiltonian ODEs using Lie group methods, we need a group structure on the cotangent bundle of G, as well as the map f that corresponds to this type of ODEs.

1.3 Hamilton–Pontryagin Mechanics

Lagrangian mechanics on G is formulated in terms of a Lagrangian \( L: TG \to \mathbb{R} \). Hamilton’s principle states that the dynamics is given by the curve \( q: \mathbb{R} \to G \) that extremizes the action integral
\[ S_H = \int_0^T L(q, \dot{q}) \, dt, \]
where the end points \( q(0) \) and \( q(T) \) are kept fixed. In [7, Theorem 3.4], it was shown that this is equivalent to the Hamilton–Pontryagin (HP) principle, which states that the dynamics is given by extremizing
\[ S_{HP} = \int_0^T (L(q, v) + \langle p, \dot{q} - v \rangle) \, dt, \]
where \( v \in T_q G, p \in T^*_q G \) are varied arbitrarily, and the end points of \( q \) are kept fixed. Here, we denote the natural pairing of covectors and vectors by \( \langle \cdot, \cdot \rangle \). This action integral leads to dynamics formulated on \( T^* G \).

To simplify further calculations, it is convenient to right-trivialize \( T^* G \) to \( G \times \mathfrak{g}^* \) via the map \( (q, p_q) \mapsto (q, p_q \cdot q^{-1}) \). Letting \( \ell(q, \xi) := L(q, \xi \cdot q), \xi \in \mathfrak{g} \), it is easy to show that the HP principle is equivalent to the right-trivialized HP principle, which has action integral
\[ S_{HP} = \int_0^T (\ell(q, v) + \langle p, \dot{q} - v \rangle) \, dt. \]
\[ S = \int_0^T \left( \ell(q, \xi) + \langle \mu, \dot{q} \cdot q^{-1} - \xi \rangle \right) \, dt, \]

where \( \xi : \mathbb{R} \to g \) and \( \mu : \mathbb{R} \to g^* \) are varied arbitrarily, and the end points of \( q \) are kept fixed. Taking the variation of \( S \), we arrive at the right-trivialized HP equations

\begin{align*}
\dot{q} &= \xi \cdot q, \\
\dot{\mu} &= -\text{ad}_\xi^* \mu + (D_1 \ell(q, \xi)) \cdot q^{-1}, \\
\mu &= D_2 \ell(q, \xi),
\end{align*}

(2)

where \( \text{ad}_x \) is the derivative of \( \text{Ad}_{\exp(x)} \) with respect to \( x \) at the origin, and \( D_k \ell \) denotes the partial derivative of \( \ell \) with respect to the \( k \)th variable, i.e., a one form. This is the ODE on \( G \times g^* \) that we need to solve.

### 1.4 Variational Integrators

Variational integrators are constructed by discretizing an action integral and then performing extremization with fixed end points. This procedure turns the action integral into an action sum. The discrete Lagrangian \( L_h : G \times G \to \mathbb{R} \) is an approximation of the action integral over a small time step \( h \),

\[ L_h(q_{k-1}, q_k) \approx \int_{(k-1)h}^{kh} L(q, \dot{q}) \, dt, \]

where \( q : \mathbb{R} \to G \) extremizes the action integral with \( q(0) \) and \( q(T) \) fixed. Letting \( N = T / h \), the action sum becomes

\[ S_h = \sum_{k=1}^{N} L_h(q_{k-1}, q_k). \]

Extremizing \( S_h \) while keeping \( q_0 \) and \( q_N \) fixed gives us the discrete Euler–Lagrange equations

\[ D_1 L_h(q_k, q_{k+1}) + D_2 L_h(q_{k-1}, q_k) = 0, \quad 1 \leq k < N. \]

The discrete Legendre transforms define the discrete conjugate momenta

\[ p_k := \mu_k \cdot q_k := -D_1 L_h(q_k, q_{k+1}), \]

\[ p_{k+1} := \mu_{k+1} \cdot q_{k+1} := D_2 L_h(q_k, q_{k+1}). \]

By demanding that these two definitions are consistent, we automatically satisfy the discrete Euler–Lagrange equations. If we can solve the first equation for \( q_{k+1} \), we can use the second one to calculate \( \mu_{k+1} \), giving us the variational integrator \( (q_k, \mu_k) \mapsto (q_{k+1}, \mu_{k+1}) \).
From a Lie Group Method to a Variational Integrator on the Cotangent Bundle

2.1 Group Structure and Hamiltonian ODEs on $G \times g^*$

We want to numerically solve the right-trivialized HP Eq. (2), which can be viewed as a vector field on $G \times g^*$, or equivalently, as the ODE $\dot{z} = f(z) \cdot z$, where $z \in G \times g^*$ and $f: G \times g^* \to g \times g^*$. For this ODE to make sense, we must choose a group product on $G \times g^*$. We choose the magnetic extension of $G$, as described by Arnold and Khesin [11, Section I.10.B]. As we will see, this group product makes the right-trivialized HP equations easily expressible as $\dot{z} = f(z) \cdot z$.

The magnetic extension assigns the following group product to $T^*G$:

$$ (g, p_g)(h, p_h) := (gh, p_g \cdot h + g \cdot p_h). \quad (3) $$

This group structure is an extension of the group structure on $G$ in the sense that the canonical projection $T^*G \to G$ is a homomorphism of Lie groups.

We note that $(g, p_g)(h, p_h) = (gh, p_g \cdot h + g \cdot p_h) = \left( gh, \left( p_g \cdot g^{-1} + Ad_{g^{-1}}^* (p_h \cdot h^{-1}) \right) \cdot gh \right)$. Thus, letting $\mu = p_g \cdot g^{-1}$ and $v = p_h \cdot h^{-1}$, the right-trivialized version of (3) is the product on $G \times g^*$ defined by

$$(g, \mu)(h, v) := \left( gh, \mu + Ad_{g^{-1}}^* v \right).$$

It can be shown that the Lie algebra associated with the Lie group $G \times g^*$ is $g \times g^*$ equipped with the Lie bracket $[(\xi, \mu), (\eta, v)] = (\text{ad}_{\xi}^* \eta, \text{ad}_{\eta}^* \mu - \text{ad}_{\eta}^* v)$. We will also need an expression for $TR_z \xi$ for $z = (q, \mu) \in \tilde{G} \times g^*$ and $\xi = (\eta, v) \in g \times g^*$:

$$ TR_z \xi = \left. \frac{d}{d\epsilon} \left( \exp(\epsilon \eta), \epsilon v \right) (q, \mu) \right|_{\epsilon=0} = \left. \frac{d}{d\epsilon} \left( \exp(\epsilon \eta)q, \epsilon v + \text{Ad}_{\exp(-\epsilon \eta)}^* \mu \right) \right|_{\epsilon=0} = \left( \eta \cdot q, v - \text{ad}_{\eta}^* \mu \right).$$

We would now like to use this to write the right-trivialized HP Eq. (2) in the form of (1), $\dot{z} = f(z) \cdot z = TR_z \circ f(z)$. If the map $f: G \times g^* \to g \times g^*$ satisfies

---

1 This group structure was used by Engø in [12] to construct partitioned Runge–Kutta–Munthe-Kaas methods on $T^*G$, without any special regard to symplecticity.
for all \((q, \xi) \in G \times g\), we see that \(\dot{z} = f(z) \cdot z\), which is exactly what we need.

In many cases, the map \((q, \xi) \mapsto (q, D_2 \ell(q, \xi))\) is a diffeomorphism of manifolds. If this holds, we say that the Lagrangian \(\ell\) is \emph{regular}. If \(\ell\) is regular, the Lagrangian problem has an equivalent formulation as a Hamiltonian ODE on \(T^*G\), where

\[
\mathcal{H}(q, D_2 \ell(q, \xi)) = (D_2 \ell(q, \xi), \xi) - \ell(q, \xi)
\]

and

\[
f(q, \mu) = \left( D_2 \mathcal{H}(q, \mu), -(D_1 \mathcal{H}(q, \mu)) \cdot q^{-1} \right).
\]

For Hamiltonians which arise in this manner, the map \((q, \mu) \mapsto (q, D_2 \mathcal{H}(q, \mu))\) is also a diffeomorphism of manifolds. In fact, the map is the inverse of the one above. Hamiltonians for which this hold are also called \emph{regular}.

\section{2.2 General Format for Our Integrators}

It is natural to consider discrete Lagrangians based on the approximation of the action integral by quadrature. The procedure adopted in the present article is inspired by the approach in [4, Section VI.6.3], originally found in [13]. In this reference, the symplectic partitioned Runge–Kutta methods are derived by considering the discrete Lagrangian

\[
L_h(q_0, q_1) = h \sum_{i=1}^{s} b_i L(Q_i, \dot{Q}_i)
\]

where

\[
Q_i = q_0 + h \sum_{j=1}^{s} a_{ij} \dot{Q}_j,
\]

and \(b_i, a_{ij}\) is the coefficients of a Runge–Kutta method. The \(\dot{Q}_i\) is chosen to extremize the sum above under the constraint

\[
q_1 = q_0 + h \sum_{i=1}^{s} b_i \dot{Q}_i.
\]

As shown in [4, Section VI.6.3], the resulting integrator is exactly the partitioned Runge–Kutta integrator where the position is integrated using the original coefficients \(b_i, a_{ij}\), while the momentum is integrated by using the coefficients \(\hat{b}_i = b_i, \hat{a}_{ij} = b_j - b_j a_{ji} / b_i\).
In the following, we will generalize the approach used in [4, Section VI.6.3] to Lie groups. Consider the discrete Lagrangian

\[ L_h(q_0, q_1) = \hat{L}_h(Q_1, \ldots, Q_s, \xi_1, \ldots, \xi_s) = h \sum_{i=1}^{s} b_i \ell(Q_i, \xi_i), \]

where \( b_i \) are nonzero quadrature weights, and the auxiliary variables \( Q_1, \ldots, Q_s, \xi_1, \ldots, \xi_s \) are chosen to extremize \( \hat{L}_h \) under the constraints

\[
Y(Q_1, \ldots, Q_s, \xi_1, \ldots, \xi_s, q_0) - \log \left( \frac{q_1}{q_0} \right) = 0,
\]

\[
X_i(Q_1, \ldots, Q_s, \xi_1, \ldots, \xi_s, q_0) - \log \left( \frac{Q_i}{q_0} \right) = 0, \quad i = 1, \ldots, s.
\]

(7)

The functions \( Y \) and \( X_i \) will typically arise from Lie group integrators, as we will see later on. The formulation of the discrete Lagrangian is that of a constrained optimization problem. As done in [4, Section VI.6.3], we solve this by introducing Lagrange multipliers. Let \( \Lambda \) be the Lagrange multiplier corresponding to the constraint containing \( Y \), and let \( \lambda_i \) be the Lagrange multiplier corresponding to the equation containing \( X_i \) for \( i = 1, \ldots, s \). To obtain a variational integrator, we extremize

\[
\hat{L}_h - \left( \Lambda, Y - \log \left( \frac{q_1}{q_0} \right) \right) - \sum_{i=1}^{s} \left( \lambda_i, X_i - \log \left( \frac{Q_i}{q_0} \right) \right),
\]

while keeping \( q_0 \) and \( q_1 \) fixed. Varying this with respect to \( \Lambda, \lambda_i, \xi_i, \) and \( Q_i \), we obtain the set of equations

\[
q_1 = \exp(Y)q_0,
\]

\[
Q_i = \exp(X_i)q_0,
\]

\[
\frac{\partial \hat{L}_h}{\partial \xi_i} = \left( \frac{\partial Y}{\partial \xi_i} \right)^* \Lambda + \sum_j \left( \frac{\partial X_j}{\partial \xi_i} \right)^* \lambda_j,
\]

\[
\frac{\partial \hat{L}_h}{\partial Q_i} = \left( \frac{\partial Y}{\partial Q_i} \right)^* \Lambda + \sum_j \left( \frac{\partial X_j}{\partial Q_i} \right)^* \lambda_j - \left( \left( \text{dexp}_{X_i}^{-1} \right)^* \lambda_i \right) \cdot Q_i,
\]

for all \( i = 1, \ldots, s \).

To find the integrator based on the discrete Lagrangian \( L_h \), we need to evaluate the partial derivatives of \( L_h \) with respect to \( q_0 \) and \( q_1 \). In doing so, we consider \( Q_1, \ldots, Q_s \) and \( \xi_1, \ldots, \xi_s \) as functions of \( q_0 \) and \( q_1 \) defined implicitly by (7) and (8). The partial
derivatives of $L_h$ are then

\[
\frac{\partial L_h}{\partial q_0} = \sum_j \left( \frac{\partial \hat{L}_h}{\partial Q_j} \circ \frac{\partial q_j}{\partial q_0} + \frac{\partial \hat{L}_h}{\partial \xi_j} \circ \frac{\partial \xi_j}{\partial q_0} \right),
\]

\[
\frac{\partial L_h}{\partial q_1} = \sum_j \left( \frac{\partial \hat{L}_h}{\partial Q_j} \circ \frac{\partial q_j}{\partial q_1} + \frac{\partial \hat{L}_h}{\partial \xi_j} \circ \frac{\partial \xi_j}{\partial q_1} \right).\]

(9)

The functions $Q_1, \ldots, Q_s, \xi_1, \ldots, \xi_s$ satisfy the constraints (7) for all $q_0, q_1$. By differentiating the constraints, we see that the identities

\[
0 = \frac{\partial Y}{\partial q_0} + \sum_j \left( \frac{\partial Y}{\partial Q_j} \circ \frac{\partial q_j}{\partial q_0} + \frac{\partial Y}{\partial \xi_j} \circ \frac{\partial \xi_j}{\partial q_0} \right) + \text{dexp}_{-Y}^{-1} \circ TR_{q_0^{-1}},
\]

\[
0 = \sum_j \left( \frac{\partial Y}{\partial Q_j} \circ \frac{\partial q_j}{\partial q_1} + \frac{\partial Y}{\partial \xi_j} \circ \frac{\partial \xi_j}{\partial q_1} \right) - \text{dexp}_Y^{-1} \circ TR_{q_1^{-1}}.
\]

\[
0 = \frac{\partial X_i}{\partial q_0} + \sum_j \left( \frac{\partial X_i}{\partial Q_j} \circ \frac{\partial q_j}{\partial q_0} + \frac{\partial X_i}{\partial \xi_j} \circ \frac{\partial \xi_j}{\partial q_0} \right) + \text{dexp}_{-X_i}^{-1} \circ TR_{Q_{i}^{-1}} - \text{dexp}_{X_i}^{-1} \circ TR_{Q_{i}^{-1}} \circ \frac{\partial Q_i}{\partial q_0},
\]

\[
0 = \sum_j \left( \frac{\partial X_i}{\partial Q_j} \circ \frac{\partial q_j}{\partial q_1} + \frac{\partial X_i}{\partial \xi_j} \circ \frac{\partial \xi_j}{\partial q_1} \right) - \text{dexp}_{X_i}^{-1} \circ TR_{Q_{i}^{-1}} \circ \frac{\partial Q_i}{\partial q_1}, \quad i = 1, \ldots, s,
\]

(10)

all hold.

We combine the discrete Legendre transforms

\[
-\mu_0 \cdot q_0 = \frac{\partial L_h}{\partial q_0}, \quad \mu_1 \cdot q_1 = \frac{\partial L_h}{\partial q_1},
\]

(11)

with (8) and (9), and simplify using (10) to obtain the equations

\[
\mu_0 = \left( \left( \frac{\partial Y}{\partial q_0} \right)^* \Lambda + \sum_j \left( \frac{\partial X_i}{\partial q_0} \right)^* \lambda_j \right) \cdot q_0^{-1} + \left( \text{dexp}_{-Y}^{-1} \right)^* \Lambda + \sum_j \left( \text{dexp}_{-X_j}^{-1} \right)^* \lambda_j,
\]

\[
\mu_1 = \left( \text{dexp}_Y^{-1} \right)^* \Lambda.
\]

Using the identity (4), we get

\[
f(Q_i, D_2 \ell(Q_i, \xi_i)) = \left( \xi_i, (D_1 \ell(Q_i, \xi_i)) \cdot Q_i^{-1} \right),
\]
and defining \( n_i, M_i \in \mathfrak{g}^* \) by

\[
\frac{\partial \hat{L}_h}{\partial Q_i} = h b_i \mathbf{D}_1 \ell(Q_i, \xi_i) = h b_i n_i \cdot Q_i, \quad \frac{\partial \hat{L}_h}{\partial \xi_i} = h b_i \mathbf{D}_2 \ell(Q_i, \xi_i) = h b_i M_i,
\]

for \( i = 1, \ldots, s \), we get

\[
h b_i n_i = \left( \left( \frac{\partial Y}{\partial Q_i} \right)^* + \sum_j \left( \frac{\partial X_j}{\partial Q_i} \right)^* \lambda_j \right) \cdot Q_i^{-1} - \left( \text{dexp}_{X_i}^{-1} \right)^* \lambda_i,
\]

\[
h b_i M_i = \left( \frac{\partial Y}{\partial \xi_i} \right)^* + \sum_j \left( \frac{\partial X_j}{\partial \xi_i} \right)^* \lambda_j.
\]

Combining everything above, the variational integrator is defined by the set of equations

\[
\begin{align*}
\mu_0 &= \left( \left( \frac{\partial Y}{\partial q_0} \right)^* + \sum_j \left( \frac{\partial X_j}{\partial q_0} \right)^* \lambda_j \right) \cdot q_0^{-1} + \left( \text{dexp}_{-Y}^{-1} \right)^* \Lambda + \sum_j \left( \text{dexp}_{-X_j}^{-1} \right)^* \lambda_j, \\
h b_i n_i &= \left( \left( \frac{\partial Y}{\partial Q_i} \right)^* + \sum_j \left( \frac{\partial X_j}{\partial Q_i} \right)^* \lambda_j \right) \cdot Q_i^{-1} - \left( \text{dexp}_{X_i}^{-1} \right)^* \lambda_i, \\
h b_i M_i &= \left( \frac{\partial Y}{\partial \xi_i} \right)^* + \sum_j \left( \frac{\partial X_j}{\partial \xi_i} \right)^* \lambda_j, \\
(\xi_i, n_i) &= f(Q_i, M_i), \\
Q_i &= \exp(X_i) q_0, \quad i = 1, \ldots, s, \\
q_1 &= \exp(Y) q_0, \\
\mu_1 &= \left( \text{dexp}_{Y}^{-1} \right)^* \Lambda.
\end{align*}
\]

Notice that we no longer involve the Lagrangian. We only need to evaluate the vector field through the map \( f \). This opens up the possibility of applying the method to degenerate Hamiltonian systems (or indeed to any ODE on \( T^*G \)).

It should be noted that since the integrator can be formulated as a variational integrator on \( G \), the group structure chosen for \( T^*G \) in (3) is not consequential. Indeed, the integrator, as a one-step method on \( T^*G \), is uniquely defined by (7), (8), (9), and the discrete Legendre transforms (11). Neither of these depend on the introduced group structure on \( T^*G \), only on the group structure on \( G \). The group structure on \( T^*G \) is used to express (7), (8), (9), (11) as the set of Eq. (12). Note that the group structure is also involved in the definition of \( f \).

---

2 Variational methods for degenerate Hamiltonian systems using Type II generating functions have been proposed by Leok and Zhang [14].
For any choice of Lie group structure on $T^*G$, such that the canonical projection $T^*G \to G$ is a homomorphism of Lie groups, we could have started with (7), (8), (9), (11) and performed similar manipulations to obtain a set of equations describing the integrator. The set of equations so obtained would be different from (12), but the solutions $(q_1, \mu_1)$ would be the same.

### 3 First- and Second-Order Integrators

In the article by Celledoni et al. [8], a special case of variational integrators of the form introduced in the previous section was considered. These integrators serve as an example of application of the formulae above. In these methods, let $a_{ij}$ and $b_i$ be the coefficients of an $s$-stage Runge–Kutta method which satisfies $b_i \neq 0$ for all $i$. Let the discrete Lagrangian be given by

$$L_h(q_0, q_1) = h \sum_{i=1}^{s} b_i \ell(Q_i, \xi_i),$$

and the constraints by (7) and

$$Y = h \sum_{i=1}^{s} b_i \xi_i, \quad X_i = h \sum_{j=1}^{s} a_{ij} \xi_j, \quad i = 1, \ldots, s.$$

We can see that for $i, j = 1, \ldots, s$,

$$\frac{\partial Y}{\partial q_0} = 0, \quad \frac{\partial X_j}{\partial q_0} = 0,$$
$$\frac{\partial Y}{\partial Q_i} = 0, \quad \frac{\partial X_j}{\partial Q_i} = 0,$$
$$\frac{\partial Y}{\partial \xi_i} = h b_i, \quad \frac{\partial X_j}{\partial \xi_i} = h a_{ji}.$$

By inserting these into (12), we get the set of equations

$$\mu_0 = \left( \text{dexp}_{-Y}^{-1} \right)^* \Lambda + \sum_j \left( \text{dexp}_{-X_j}^{-1} \right)^* \lambda_j,$$
$$h b_i n_i = - \left( \text{dexp}_{X_j}^{-1} \right)^* \lambda_i,$$
$$h b_i M_i = h b_i \Lambda + \sum_j h a_{ji} \lambda_j,$$
$$(\xi_i, n_i) = f(Q_i, M_i),$n_i = f(Q_i, M_i),$$
$$Q_i = \exp(X_i) q_0, \quad i = 1, \ldots, s,$$
$$q_1 = \exp(Y) q_0,$$
\[ \mu_1 = \left( \text{dexp}_{\mathbf{Y}}^{-1} \right)^* \Lambda. \]

In these equations, \( \Lambda \) and \( \lambda_j \) can be eliminated, giving the integrator

\[ b_i M_i = b_i \text{dexp}^*_{\mathbf{Y}} \left( \mu_0 + h \sum_j b_j \text{Ad}_{\text{exp}(\mathbf{X}_j)}^* n_j \right) - h \sum_j b_j a_{ji} \text{dexp}^*_{\mathbf{X}_j} n_j, \]

\[ X_i = h \sum_j a_{ij} \xi_j, \]

\[ Q_i = \exp(\mathbf{X}_i) q_0, \]

\[ (\xi_i, n_i) = f(\mathbf{Q}_i, M_i), \quad i = 1, \ldots, s, \]

\[ Y = h \sum_j b_j \xi_j, \]

\[ q_1 = \exp(Y) q_0, \]

\[ \mu_1 = \text{Ad}_{\text{exp}(-Y)}^* \left( \mu_0 + h \sum_j b_j \text{Ad}_{\text{exp}(\mathbf{X}_j)}^* n_j \right). \]

Here, we have used the identity \( \text{dexp}_x \circ \text{dexp}_{x^{-1}} = \text{Ad}_{\text{exp}(x)} \). Equation (13) is equivalent to the method presented in [8, Section 5].

Methods of this form suffer from an order barrier. They cannot obtain higher accuracy than second order. The proof, presented below, is closely related to a similar order barrier for commutator-free Lie algebra methods [15].

**Proposition 3.1** The integrators of the format (13) cannot achieve higher than second order on general Hamiltonian differential equations.

**Proof** The proof proceeds by applying the variational method (13) to a particular class of regular Hamiltonian problems and a particular choice of starting values. We show that in this case, the Lie group part of the solution has an error of at most second order; thus, the variational method is at most of second order as well.

Let a Hamiltonian on \( G \times \mathfrak{g}^* \) be given by \( \mathcal{H}(q, \mu) = \langle \mu, v(q) \rangle + T(\mu) \), where \( v: G \to \mathfrak{g} \) is smooth, but otherwise arbitrary, and \( T: \mathfrak{g}^* \to \mathbb{R} \) is a non-degenerate quadratic function of \( \mu \). Using (5), we find that the corresponding Hamiltonian vector field is

\[ f(q, \mu) = \left( v(q) + \frac{dT}{d\mu}, -\left( \left( \frac{\partial v}{\partial q} \right)^* \mu \right) \cdot q^{-1} \right), \]

and the differential equation is

\[ \dot{q} = \left( v(q) + \frac{dT}{d\mu} \right) \cdot q, \]

\[ \dot{\mu} = -\left( \left( \frac{\partial v}{\partial q} \right)^* \mu \right) \cdot q^{-1} - \text{ad}_{v(q) + \frac{dT}{d\mu}}^\ast \mu. \]
We note that $\frac{dT}{d\mu}$ and $U(q)\mu := -\left(\frac{\partial v}{\partial q}\right)^*\mu \cdot q^{-1}$ are both linear in $\mu$, in particular $\frac{dT}{d\mu}\big|_{\mu=0} = 0$.

A particular class of solutions to this ODE consists of those solutions which satisfy $\mu(t) = 0$ for all $t$. For these solutions, $q(t)$ solves the ODE $\dot{q} = v(q) \cdot q$. We want to show that the numerical solution from (13), when applied to this problem, preserves the invariant $\mu = 0$, and that the method reduces to a conventional Lie group method in this case. If we apply (13) to the Hamiltonian vector field (14) and set $\mu_0 = 0$, we get, among others, the equation $n_i = U(Q_i)M_i$. Inserting this into the first equation of (13), we get

$$b_i M_i = h b_i \exp^*_{-Y} \sum_j b_j \text{Ad}^*_{\exp(X_j)} U(Q_j)M_j$$

$$-h \sum_j b_j a_{ji} \exp^*_{X_j} U(Q_j)M_j, \quad i = 1, \ldots, s.$$

Clearly, this system of equations has $M_i = 0$ for $i = 1, \ldots, s$ as a solution. Additionally, if we assume that $Y$ and $X_j$ go to zero as $h$ goes to zero, $M_i = 0$ is the only solution for small enough step-length $h$. Therefore, $n_i = U(Q_i)M_i = 0$ and $\mu_1 = 0$. The remaining equations of (13) are

$$X_i = h \sum_j a_{ij} \xi_j,$$

$$Q_i = \exp(X_i)q_0,$$

$$\xi_i = v(Q_i) + \left(\frac{dT}{d\mu}\right)_{\mu=M_i} = v(Q_i), \quad i = 1, \ldots, s,$$

$$Y = h \sum_j b_j \xi_j,$$

$$q_1 = \exp(Y)q_0.$$

We recognize these equations as a commutator-free Lie group method with one exponential, or equivalently, an RKMK method with cutoff parameter 0, applied to the ODE $\dot{q} = v(q) \cdot q$. As explained in [16], commutator-free methods with one exponential cannot satisfy the third-order conditions, and the solution is at most second-order accurate.

By repeating the argument with any variational integrator of the form described in (12), we get a generalization.

**Proposition 3.2** A variational integrator on $T^*G$ of the form (12) based on a Lie group integrator cannot achieve higher order than the underlying Lie group integrator.

We should note that first- and second-order methods of the format described in (13) do exist. Specifically, a method of that format is first order if $\sum_{i=1}^s b_i = 1$ and second
order if, in addition, \( \sum_{i,j=1}^{s} b_i a_{ij} = 1/2 \). The proof is a special case of Theorem 5.3 with cutoff parameter \( r = 0 \).

**Example 3.3** *(The midpoint method)* Let us choose \( s = 1 \). The method is of second order if and only if we choose \( b_1 = 1 \) and \( a_{11} = 1/2 \). This method is also symmetric, since the substitutions \( h \to -h, q_0 \leftrightarrow q_1, \mu_0 \leftrightarrow \mu_1, Y \to -Y \) in (13) yields the same method after some manipulation of the equations. This property is utilized in Sect. 6.1 to achieve high order by composition.

4 Higher-Order Integrators

The methods discussed in the previous section were limited to at most second order. To obtain higher-order integrators, we consider two approaches, based on two well-known classes of Lie group integrators.

The first approach is based on the Runge–Kutta–Munthe-Kaas (RKMK) methods. This approach was already considered by Bou-Rabee and Marsden in 2009 [7]. The work in the present article builds on the work by Bou-Rabee and Marsden and examines in detail the case when the cutoff parameter \( r \) (\( q \) in [2,7]) in the RKMK method is larger than 0 and provides a complete order theory for variational methods based on RKMK methods.

The second approach is based on Crouch–Grossman (CG) methods. This approach has to our knowledge not been explored before. We show that these methods can achieve arbitrarily high order, but the complete order theory of these methods remains unresolved.

4.1 Variational Runge–Kutta–Munthe-Kaas Integrators

A popular class of Lie group integrators is the class of RKMK integrators. For our purposes, these integrators can be written

\[
\begin{align*}
  x_i &= h \sum_{j=1}^{s} a_{ij} \text{dexp}_{(r),x_j}^{-1} \xi_j, \\
  Q_i &= \exp(x_i) q_0, \\
  \xi_i &= f(Q_i), \quad i = 1, \ldots, s, \\
  Y &= h \sum_{i=1}^{s} b_i \text{dexp}_{(r),x_i}^{-1} \xi_i, \\
  q_1 &= \exp(Y) q_0,
\end{align*}
\]

where

\[
\text{dexp}_{(r),x}^{-1} = \text{id} - \frac{1}{2} \text{ad}_x + \sum_{k=2}^{r} \frac{B_k}{k!} (\text{ad}_x)^k
\]
is the Taylor series approximation to $\text{dexp}^{-1}$, and $a_{ij}, b_i$ are the coefficients of a Runge–Kutta method. If the RK method is of order $p$ and $r \geq p - 2$, the resulting Lie group integrator is of order $p$ as well [4, Theorem IV.8.4].

Variational methods based on RKMK methods were considered by Bou-Rabee and Marsden in [7], though the methods they present in detail are at most second order, since they only consider the case $r = 0$. The methods in this case are essentially the methods considered in Sect. 3. For $r > 0$, some complications arise for the variational integrator, since the $x_i$ are not explicitly given by $\xi_1, \ldots, \xi_s$. Our solution is to treat both $x_i$ and $\xi_i$ as unknowns and the equations for $x_i$ in (15) as restrictions. The Lagrange multipliers $\lambda_i$ corresponding to the equations for $x_i$ cannot be eliminated from the equations in a general manner, so the dimension of the nonlinear equation to be solved at each step is larger than that of the corresponding symplectic method applied to $T^*\mathbb{R}^n$, or the simpler integrator with $r = 0$.

In the variational integrator, we set the discrete Lagrangian to

$$L_h = h \sum_{j=1}^{s} b_j \ell(Q_j, \xi_j),$$

and let the constraints be given by (7), that is,

$$q_1 = \exp(Y)q_0, \quad Q_i = \exp(X_i)q_0.$$

and

$$Y = h \sum_{j=1}^{s} b_j \text{dexp}^{-1}_{(r), x_j} \xi_j,$$

$$X_i = h \sum_{j=1}^{s} a_{ij} \text{dexp}^{-1}_{(r), x_j} \xi_j, \quad i = 1, \ldots, s,$$

where $x_i = \log \left( Q_i q_0^{-1} \right)$. Note that on the solution set of the constraints, $X_i = x_i$.

Applying the variational equations from (12), the integrator is given by

$$\mu_0 = \left( \left( \text{dexp}^{-1}_Y \right)^* - h \sum_i b_i \left( \text{dexp}^{-1}_{-X_i} \right)^* \circ P^*_r (X_i, \xi_i) \right) \Lambda$$

$$+ \sum_j \left( \left( \text{dexp}^{-1}_{-X_j} \right)^* - h \sum_i a_{ji} \left( \text{dexp}^{-1}_{-X_i} \right)^* \circ P^*_r (X_i, \xi_i) \right) \lambda_j,$$

$$hb_i n_i = - \left( \text{dexp}^{-1}_{X_i} \right)^* \lambda_i + hb_i \left( \text{dexp}^{-1}_{X_i} \right)^* \circ P^*_r (X_i, \xi_i) \Lambda.$$
\[ +h \sum_j a_{ji} \left( \text{dexp}_{X_i}^{-1} \right)^* \circ P^* (X_i, \xi_i) \lambda_j, \]

\[ h b_i M_i = h \left( \text{dexp}_{(r),X_i}^{-1} \right)^* \left( b_i \Lambda + \sum_j a_{ji} \lambda_j \right), \]

\[ Q_i = \exp(X_i)q_0, \]

\[ (\xi_i, n_i) = f(Q_i, M_i), \quad i = 1, \ldots, s, \]

\[ q_1 = \exp(Y)q_0, \]

\[ \mu_1 = \left( \text{dexp}_Y^{-1} \right)^* \Lambda, \quad (16) \]

where \( P^*_{(r)}(x, \xi) \) is a polynomial in \( \text{ad}^* x \) and \( \text{ad}^* \xi \) of degree \( r \), defined as the adjoint of the partial derivative of \( \text{dexp}^{-1} \) \( x \) \( \xi \) with respect to \( x \). Specifically,

\[ P^*_{(0)}(x, \xi) = 0, \]
\[ P^*_{(1)}(x, \xi) = \frac{1}{2} \text{ad}^*_\xi, \]
\[ P^*_{(2)}(x, \xi) = \frac{1}{2} \text{ad}^*_\xi - \frac{1}{6} \text{ad}^*_\xi \text{ad}^*_x + \frac{1}{12} \text{ad}^*_x \text{ad}^*_\xi, \]
\[ P^*_{(r)}(x, \xi) = \frac{1}{2} \text{ad}^*_\xi - \sum_{k=2}^{r} \frac{B_k}{k!} \sum_{i=0}^{k-1} \text{ad}^*_\xi \text{ad}^*_x \left( \text{ad}^*_x \right)^{k-i-1}. \]

By applying \( \text{Ad}^*_{\exp(X_i)} \) to both sides of the second equation in (16), we see that the first equation can be simplified. Using this and rearranging the rest of the equations while assuming that \( b_i \neq 0 \), we arrive at the set of equations

\[ \Lambda = \text{dexp}^*_Y \left( \mu_0 + h \sum_i b_i \text{Ad}^*_{\exp(X_i)} n_i \right), \]

\[ \lambda_i = -h b_i \text{dexp}^*_X n_i + h P^*_{(r)}(X_i, \xi_i) \left( b_i \Lambda + \sum_j a_{ji} \lambda_j \right), \]

\[ M_i = \frac{1}{b_i} \left( \text{dexp}_{(r),X_i}^{-1} \right)^* \left( b_i \Lambda + \sum_j a_{ji} \lambda_j \right), \]

\[ X_i = h \sum_j \sum_{i,j} a_{ij} \text{dexp}^{-1}_{(r),X_j} \xi_j, \]

\[ (\xi_i, n_i) = f(\exp(X_i)q_0, M_i), \quad i = 1, \ldots, s, \]

\[ Y = h \sum_i b_i \text{dexp}^{-1}_{(r),X_i} \xi_i, \]

\[ q_1 = \exp(Y)q_0. \]
\[
\mu_1 = \text{Ad}_{\exp(-Y)}^* \left( \mu_0 + h \sum_i b_i \text{Ad}_{\exp(X_i)}^* n_i \right),
\]  

which define a symplectic integrator on \( T^*G \). The identity \( \text{dexp}_x \circ \text{dexp}^{-1}_x = \text{Ad}_{\exp(x)} \) was used to obtain the last line of the equations above. We call the integrators defined by (17) variational Runge–Kutta–Munthe-Kaas methods or VRKMK methods for short. One can easily check that if the Lie group is abelian, a VRKMK method simplifies to a symplectic partitioned Runge–Kutta method.

In implementations of these methods, it is required that \( \exp: \mathfrak{g} \to G \) and \( \text{dexp}^*: \mathfrak{g} \times \mathfrak{g}^* \to \mathfrak{g}^* \) are calculated to machine precision to obtain symplecticity. In the numerical tests of Sect. 6, we choose \( G = \text{SO}(3) \) and use Rodrigues’ formula [17, Section 9.2] to calculate these expressions. In a general setting, calculating these expressions usually involves analytic functions of matrices. The equations defining the integrator can be solved as a set of nonlinear equations in the unknowns \( X_i, M_i \) and \( \lambda_i, i = 1, \ldots, s \), as the other quantities in (17) are given explicitly in terms of the aforementioned variables as well as \( q_0 \) and \( \mu_0 \). If \( G \) is an \( n \)-dimensional Lie group, this is a total of \( 3ns \) scalar unknowns. Other choices of independent and dependent unknowns are possible. One could reduce the number of unknowns by starting with coefficients of an explicit RK method and thereby obtaining explicit expressions for \( x_i \) and be able to eliminate the \( \lambda_i \) in the integrator. The variational method based on an explicit RK method would still be implicit, however, and due to the order conditions presented later in this article, the increased number of stages required for a particular order would offset the reduction in number of unknowns per stage obtained by using an explicit method as the underlying method, so it is unclear if this simplification is useful in practice.

4.1.1 Alternative Approach

The authors have discovered an alternative approach which reduces the number of unknowns in the Eq. (17). The alternative approach requires a modification of the variational principle described in Sect. 2, and full details and analysis goes beyond the scope of this article.

A modification of the RKMK methods can be obtained by replacing the Taylor approximation \( \text{dexp}_r^{-1}(r) \) in (15) with a Padé approximation. After trivial manipulations, the modified RKMK method is

\[
\begin{align*}
x_i &= h \sum_j a_{ij} \tilde{\xi}_j, \\
Q_i &= \exp(x_i)q_0, \\
\text{dexp}_{(r),x_i} \tilde{\xi}_i &= f(Q_i), \quad i = 1, \ldots, s, \\
Y &= h \sum_i b_i \tilde{\xi}_i, \\
q_1 &= \exp(Y)q_0,
\end{align*}
\]
where $\exp_{(r,\tilde{\xi})} = 1 + \frac{1}{2} \text{ad}_{\tilde{\xi}} + \cdots + \frac{1}{(r+1)!} \text{ad}_{\tilde{\xi}}^r$. In this formulation, the $x_j$ are explicit in the $\tilde{\xi}_j$, so there is no need to introduce restrictions for the equations $x_i = \sum_j a_{ij} \tilde{\xi}_j$.

The discrete Lagrangian in this formulation is

$$L_h(q_0, q_1) = h \sum_j b_j \ell(Q_j, \exp_{(r, x_j)} \tilde{\xi}_j),$$

which is not of the format (6) discussed in Sect. 2. Therefore, the general formulae for integrators (12) derived earlier do not apply. However, the general idea can still be pursued, and the resulting integrator can be formulated on the Hamiltonian side.

4.2 Variational Crouch–Grossman Integrators

The methods of Crouch and Grossman form another important class of Lie group methods. Crouch and Grossman formulated their integrators in terms of rigid frames, i.e., finite collections of vector fields on a manifold. On a Lie group, a suitable rigid frame is a basis for the right-invariant vector fields on $G$ corresponding to a basis of $\mathfrak{g}$. In this setting, the Crouch–Grossman methods can be defined as follows. Let $b_i$, $a_{ij}$ be the coefficients of an $s$-stage RK method. The Crouch–Grossman method [4, Section IV.8.1] with the same coefficients is defined by the equations

$$Q_i = \exp(ha_{i,s} \xi_s) \cdots \exp(ha_{i,1} \xi_1) q_0,$$

$$\tilde{\xi}_i = f(Q_i),$$

$$q_1 = \exp(hb_s \xi_s) \cdots \exp(hb_1 \xi_1) q_0.$$ 

The order of a CG method is determined by the order conditions developed in [18].

Using the general format (12), we set the discrete Lagrangian to

$$L_h(q_0, q_1) = h \sum_{i=1}^s b_i \ell(Q_i, \tilde{\xi}_i), \quad (18)$$ 

with $b_i \neq 0$ and constraints given by (7), that is,

$$q_1 = \exp(Y) q_0,$$

$$Q_i = \exp(X_i) q_0,$$

and

$$Y = \log(\exp(hb_s \xi_s) \cdots \exp(hb_1 \xi_1)),$$

$$X_i = \log(\exp(ha_{i,s} \xi_s) \cdots \exp(ha_{i,1} \xi_1)).$$
Inserting this into the equations defining a variational integrator (12), we obtain

$$\mu_0 = \left(\exp^{-1}_{-Y}\right)^* \Lambda + \sum_j \left(\exp^{-1}_{-X_j}\right)^* \lambda_j,$$

$$hb_i n_i = -\left(\exp^{-1}_{X_i}\right)^* \lambda_i,$$

$$hb_i M_i = hb_i \exp_{h b_i \xi_i} \circ \Ad_{\exp\left(h b_s, \xi_i\right) \cdots \exp\left(h b_{i+1}, \xi_{i+1}\right)} \circ \left(\exp^{-1}_Y\right)^* \Lambda$$

$$+ h \sum_j a_{ji} \exp_{h a_j \xi_i} \circ \Ad_{\exp\left(h a_{js}, \xi_s\right) \cdots \exp\left(h a_{j,i+1}, \xi_{i+1}\right)} \circ \left(\exp^{-1}_{X_j}\right)^* \lambda_j,$$

$$(\xi_i, n_i) = f(Q_i, M_i),$$

$$Q_i = \exp(X_i)q_0,$$

$$q_1 = \exp(Y)q_0,$$

$$\mu_1 = \left(\exp^{-1}_Y\right)^* \Lambda.$$

Eliminating $\Lambda$ and $\lambda_j$, and rearranging, we get the integrator

$$q_1 = q^s,$$

$$q^j = \exp(h b_j \xi_j)q^{j-1},$$

$$q^0 = q_0,$$

$$(\xi_i, n_i) = f(Q_i, M_i),$$

$$\bar{\mu}_0 = \Ad_{q_0}^* \mu_0, \quad \bar{\mu}_1 = \Ad_{q_1}^* \mu_1,$$

$$\bar{n}_i = \Ad_{Q_i}^* n_i,$$

$$\bar{\mu}_1 = \bar{\mu}_0 + h \sum_{j=1}^s b_{j} \bar{n}_j,$$

$$M_i = \exp_{h b_i \xi_i} \circ \Ad_{(q^j)^{-1}} \bar{\mu}_1 - h \sum_{j=1}^s b_{j} a_{ji} \exp_{h a_j \xi_i} \circ \Ad_{(q^j)^{-1}} \bar{n}_j,$$  \hspace{1cm} (19)

for all $i = 1, \ldots, s$. We call the integrators defined by (19) variational Crouch–Grossman methods or VCG methods for short. The last equation in (19) can also be written as

$$M_i = \exp_{h b_i \xi_i} \circ \Ad_{(q^j)^{-1}} \bar{\mu}_0$$

$$+ h \sum_{j=1}^s b_{j} \left(\exp_{h b_j \xi_i} \circ \Ad_{(q^j)^{-1}} \frac{a_{ji}}{b_i} \exp_{h a_j \xi_i} \circ \Ad_{Q_j^{-1}} \bar{n}_j\right) \bar{n}_j,$$  \hspace{1cm} (20)

which we will need in the order analysis of Proposition 5.6.

In the case that the Lie group is abelian, the integrator simplifies to the same symplectic, partitioned RK method as in the abelian case for the VRKMK integrator.

### 5 Order Analysis

In analyzing the order of variational methods, we use variational error analysis as described by Marsden and West [5, Section 2.3]. We recite two definitions from this
reference, which are useful in the following sections. The exact discrete Lagrangian is given by

\[ L^E_h(q_0, q_1) = \int_0^h L(q(t), \dot{q}(t)) \, dt, \]

where \( q(t) \) is the solution to the Euler–Lagrange equations with \( q(0) = q_0, q(h) = q_1 \). A discrete Lagrangian \( L_h \) is said to be of order \( p \) if

\[ L_h(q(0), q(h)) = L^E_h(q(0), q(h)) + O(h^{p+1}), \]

for all solutions \( q(t) \) of the Euler–Lagrange equations.

The following theorem is a special case of [5, Theorem 2.3.1].

**Theorem 5.1** Given a regular Lagrangian \( L \) and a discrete Lagrangian \( L_h \) of order \( p \), then the symplectic integrator defined by \( L_h \) is of order \( p \).

Both classes of methods presented in this article depend on Butcher coefficients \( a_{ij} \) and \( b_i \). Furthermore, when applied to an abelian Lie group (for instance \( \mathbb{R}^n \)), both classes become symplectic, partitioned RK methods where the position is integrated with the RK method with coefficients \( a_{ij} \) and \( b_i \), while the momentum is integrated with the RK method with coefficients \( \hat{a}_{ij} = b_j - b_j a_{ji} / b_i \) and \( \hat{b}_i = b_i \). The order conditions for SPRK methods have been explored in detail by Murua [20]. Since an abelian Lie group is a special case, the order of the SPRK method is an upper bound for the order of the variational Lie group method with the same coefficients. The order of the underlying Lie group method is also an upper bound according to Proposition 3.2.

### 5.1 Order of VRKMK Integrators

The VRKMK methods described in Sect. 4.1 are fully described by the Butcher coefficients \( a_{ij} \) and \( b_i \), and the cutoff parameter \( r \). The cutoff parameter \( r \) limits the order of the RKMK method (15) on \( G \). The order of the RKMK method is the minimum of the order of the RK method based on the same coefficients and \( r + 2 \). [4, Section IV.8.2]

As explained above, the order of the VRKMK method is bounded from above by the order of the SPRK method based on the same Butcher coefficients and by the order of the RKMK method. Since the order conditions for RK methods for a particular order form a subset of the order conditions for the SPRK method, we can a priori say that the order of the VRKMK method is bounded from above by the order of the SPRK method and \( r + 2 \). Theorem 5.3 states that in the case of a regular Lagrangian, the order of the VRKMK method is in fact the minimum of these two bounds. The proof of this theorem relies on the following lemma.

---

3 Patrick and Cuell [19] demonstrate an inaccuracy in the proof in [5]. However, they also show that the relevant result still holds.
Lemma 5.2 Assume that the continuous Lagrangian is regular, and that the SPRK method based on the coefficients $a_{ij}$ and $b_i$ is of order $d$. Then, the discrete Lagrangian of the SPRK method (6) is also of order $d$.

**Proof** The proof begins by applying the SPRK method to a Hamiltonian system, augmented with the ODE for the action integral. The order of the discrete Lagrangian is then calculated using a Taylor expansion around the solution of the one-step method.

Let $H(q, p)$ be a regular Hamiltonian, $L(q, \dot{q}) = \langle p, \dot{q} \rangle - H(q, p)$ the corresponding Lagrangian, and $(q(t), p(t))$ an exact solution to the Hamiltonian system. The resulting integrator is order $d$ accurate if and only if the original coefficients $b_i$ and $a_{ij}$ together with $\hat{b}_i = b_i$ and $\hat{a}_{ij} = b_j - b_j a_{ji}/b_i$ fulfill the order conditions up to order $d$ for a partitioned Runge–Kutta method. We will apply the partitioned Runge–Kutta method to the system

$$
\begin{bmatrix}
\dot{q} \\
\dot{\hat{S}}
\end{bmatrix} = 
\begin{bmatrix}
\frac{\partial H}{\partial p} \\
L\left(q, \dot{q}\right)
\end{bmatrix},
$$

$$
\dot{p} = -\frac{\partial H}{\partial q},
$$

where the $(q, S)$-component is integrated using the coefficients $b_i$, $a_{ij}$, and the $p$-component is integrated using the coefficients $\hat{b}_i$, $\hat{a}_{ij}$. These are simply the Hamiltonian equations augmented with the differential equation for the action integral $S$.

As starting values, we use $q_0 = q(0)$, $p_0 = p(0)$, and $S_0 = S(0) = 0$. The exact solution of the system at $t = h$ is given by the solution to the Hamiltonian equation, $q(h)$, $p(h)$, and

$$S(h) = \int_0^h L(q, \dot{q}) \, dt = L^E_h(q_0, q(h)).$$

The numerical solution obtained with one step of the partitioned Runge–Kutta method is, using the notation of Sect. 2.2,

$$q_1 = q(h) + O(h^{d+1}), \quad p_1 = p(h) + O(h^{d+1})$$

$$S_1 = h \sum_{i=1}^s b_i L(Q_i, \dot{Q}_i) = L_h(q_0, q_1) = L^E_h(q_0, q(h)) + O(h^{d+1}),$$

since the method is order $d$. We thus have a bound for the difference $L_h(q_0, q_1) - L^E_h(q_0, q(h))$. We need to establish a bound for the difference $L_h(q_0, q(h)) - L^E_h(q_0, q(h))$. We use the Taylor series expansion $L_h(q_0, q(h)) = L_h(q_0, q_1) + \langle D_2 L_h(q_0, q_1), q(h) - q_1 \rangle + O(\|q(h) - q_1\|^2)$. We then use that $p_1 = D_2 L_h(q_0, q_1)$ and the bound for $q(h) - q_1$ from (21), we see that $L_h(q_0, q(h)) - L_h(q_0, q_1) = \ldots$

---

4 Since $\hat{b}_i = b_i$ and the right-hand side is independent of $S$, we could instead have grouped $S$ with $p$ without any change.
\[ \langle p_1, q(h) - q_1 \rangle + O(\|q(h) - q_1\|^2) = O(h^{d+1}). \] Combining this bound with the last line of (21) completes the proof. □

**Theorem 5.3** If the symplectic, partitioned Runge–Kutta method based on the coefficients \( a_{ij} \) and \( b_i \) is of order at least \( p \), and the cutoff parameter \( r \) satisfies \( r \geq p - 2 \), then the variational Runge–Kutta–Munthe-Kaas method with the same coefficients is at least of order \( p \) for regular Hamiltonians.

**Proof** The proof consists of two steps. We introduce the limit case where the cutoff parameter \( r \) goes to infinity, that is, the method where \( \text{dexp}^{-1}_{(r),x} \) is replaced by \( \text{dexp}^{-1}_x \). To distinguish between the two RKMK methods, we will denote the “full” RKMK method by \( \text{RKMK}(\infty) \) and the RKMK method with cutoff parameter \( r \) by \( \text{RKMK}(r) \). The variational integrators based on the two methods are denoted as \( \text{VRKMK}(\infty) \) and \( \text{VRKMK}(r) \), respectively.

In the first step, we show that the discrete Lagrangian which defines the VRKMK(\( \infty \)) method is of order \( p \). The proof relies on two facts. Firstly, that the discrete Lagrangian of the SPRK method is of order \( p \). Secondly, that the discrete Lagrangians of the VRKMK(\( \infty \)) and of a special case of the SPRK method are obtained as extremal values of the same object function and under the same constraints.

In the second part, we show that if we apply the VRKMK(\( \infty \)) and VRKMK(\( r \)) methods to the same initial values, their difference after one step goes to zero as \( O(hr^{r+3}) \).

Let \( q: [0, a] \to G \) be a solution to the Euler–Lagrange equation with \( q(0) = q_0 \), and assume that \( a > 0 \) is sufficiently small such that \( \sigma(t) = \log(q(t)q_0^{-1}) \) is uniquely defined for all \( t \in [0, a] \). The exact discrete Lagrangian is given by

\[
L^E_h(q_0, q(h)) = \int_0^h \ell(q(t), \dot{q}(t)) \, dt, \tag{22}
\]

where \( \dot{q}(t) = \dot{q}(t) \cdot q(t)^{-1} \). If we define \( \tilde{\ell}: Tg \to \mathbb{R} \) as

\[
\tilde{\ell}(\sigma, \dot{\sigma}) = \ell(\exp(\sigma)q_0, \text{dexp}_\sigma \dot{\sigma}), \tag{23}
\]

we can rewrite (22) as \( L^E_h(q_0, q(h)) = \tilde{L}^E_h(0, \sigma(h)) \) where

\[
\tilde{L}^E_h(0, \sigma(h)) = \int_0^h \tilde{\ell}(\sigma(t), \dot{\sigma}(t)) \, dt.
\]

This is an exact discrete Lagrangian on the vector space \( g \), which we approximate by the action sum arising from the underlying RK method,

\[
\tilde{L}^\text{RK}_h(0, \sigma(h)) = h \sum_i b_i \tilde{\ell}(y_i, \eta_i), \tag{24}
\]
where \( y_i = h \sum_j a_{ij} \eta_j, i = 1, \ldots, s \) and the sum is extremized under the constraint \( \sigma(h) = h \sum_i b_i \eta_i \). Under the assumptions of the theorem, the order of the SPRK method is at least \( p \), so by Lemma 5.2, the discrete Lagrangian of the SPRK method is order \( p \) accurate,

\[
\tilde{L}_h^{\text{RK}}(0, \sigma(h)) = \tilde{L}_h^{\text{E}}(0, \sigma(h)) + O(h^{p+1}).
\]

Inserting (23) into (24) gives

\[
L_h^{\text{RK}}(q_0, q(h)) = \tilde{L}_h^{\text{RK}}(0, \sigma(h)) = h \sum_i b_i \ell \left( \exp(y_i) q_0, \text{dexp}_{y_i} \eta_i \right),
\]

where the sum is extremized under the constraint \( q(h) = \exp \left( h \sum_i b_i \eta_i \right) q_0 \).

Now, the discrete action sum arising from RKMK(\( \infty \)) is

\[
L_h^{\text{RKMK}(\infty)}(q_0, q(h)) = h \sum_i b_i \ell \left( \exp(X_i) q_0, \xi_i \right),
\]

which is extremized under the constraints

\[
X_i = h \sum_j a_{ij} \text{dexp}^{-1}_{X_j} \xi_j, \quad i = 1, \ldots, s,
q(h) = \exp \left( h \sum_i b_i \text{dexp}^{-1}_{X_i} \xi_i \right) q_0.
\]

We see that under the identifications

\[
y_i = X_i, \quad \eta_i = \text{dexp}^{-1}_{X_i} \xi_i,
\]

the objective functions (25) and (26) are identical and are extremized under the same constraints. Thus, their extremal values are identical, and we have proved

\[
L_h^{\text{RKMK}(\infty)}(q_0, q(h)) = L_h^{\text{RK}}(q_0, q(h)) = \tilde{L}_h^{\text{RK}}(0, \sigma(h)) = \tilde{L}_h^{\text{E}}(0, \sigma(h)) + O(h^{p+1}) = L_h^{\text{E}}(q_0, q(h)) + O(h^{p+1}),
\]

concluding the first part of the proof.

For the second part of the proof, we consider the integrator in (17) and the variational integrator based on RKMK(\( \infty \)) with the same initial data \( (q_0, \mu_0) \). Let \( \xi_i, n_i, X_i, M_i, \lambda_i, \Lambda \) and \( Y \) be as in (17), and \( \xi_i^{(\infty)}, \) etc., be the corresponding quantities in VRKMK(\( \infty \)).

We define \( \delta \xi_i = \xi_i - \xi_i^{(\infty)} \) and so on and consider the difference between VRKMK(\( \infty \)) and VRKMK(r). Since \( q_1 = \exp(Y) q_0 \) and \( \mu_1 = \left( \text{dexp}^{-1}_Y \right)^{*} \Lambda \), the leading order of the difference between the two integrators is given by the leading orders of \( \delta Y \) and \( \delta \Lambda \). It is clear from the equations in (17) defining the integrator
that as \( h \to 0 \), \( \lambda_i, \lambda_i^{(\infty)}, X_i, X_i^{(\infty)}, Y \) and \( Y^{(\infty)} \), all go to zero as \( \mathcal{O}(h) \). Furthermore, \( \delta \xi_i, \delta n_i, \delta M_i \) and \( \delta \Lambda \) must also go to zero as \( h \to 0 \). By using the expressions in (17), the equations \( \delta \xi_i = \delta \xi_0 + \mathcal{O}(h), n_i = n_0 + \mathcal{O}(h), \Lambda = \mu_0 + \mathcal{O}(h) \), and \( X_i = chc_i \xi_0 + \mathcal{O}(h^2) \), and the series expansions of \( \text{dexp}^* \) and \( \text{A}_{\text{dexp}^*} \), we find that

\[
\delta \Lambda = -\frac{1}{2} \text{ad}_{\delta \xi_0}^* \mu_0 + h \sum_i b_i \left( \text{ad}_{\delta \xi_i}^* n_0 + \delta n_i \right) + \text{higher-order terms},
\]

\[
\delta M_i = -\frac{1}{2} \text{ad}_{\delta \xi_i}^* \mu_0 - \frac{B_{r+1}}{(r+1)!} h^{r+1} c_i^{r+1} \left( \text{ad}_{\xi_0}^* \right)^{r+1} \mu_0 + \delta \Lambda + \sum_j \frac{a_{ij}}{b_i} \delta \lambda_j + \mathcal{O}(h^{r+2}) + \text{h.o.t.},
\]

\[
\delta \lambda_i = -h b_i \left( \frac{1}{2} \text{ad}_{\delta \xi_i}^* n_0 + \delta n_i \right) + h b_i \frac{B_{r+1}}{(r+1)!} h^r c_i^{r+1} \left( \text{ad}_{\xi_0}^* \right)^{r+1} \mu_0 + h b_i \left( \left( \frac{1}{2} \text{ad}_{\delta \xi_i}^* \frac{1}{6} \text{ad}_{\xi_0}^* \text{ad}_{\delta \xi_i}^* + \frac{1}{12} \text{ad}_{\delta \xi_i}^* \text{ad}_{\delta \xi_i}^* \text{ad}_{\xi_0}^* \right) \mu_0 + \frac{1}{2} \text{ad}_{\xi_0}^* \delta \Lambda \right) + \mathcal{O}(h^{r+2}) + \text{h.o.t.},
\]

\[
\delta X_i = h \sum_j a_{ij} \delta \xi_j + \mathcal{O}(h^{r+3}) + \text{h.o.t.},
\]

\[
(\delta \xi_i, \delta n_i) = T(q_0, \mu_0) f(\delta X_i \cdot q_0, \delta M_i) + \text{h.o.t.},
\]

\[
\delta Y = h \sum_i b_i \left( \delta \xi_i - \frac{1}{2} \text{ad}_{\delta \xi_i} \xi_0 \right) + \mathcal{O}(h^{r+3}) + \text{h.o.t.}
\]

In the equations above, “higher-order terms (h.o.t.)” denote terms that are dominated by at least one of the preceding terms.

We continue by combining the equations and dropping terms of higher order. Consider the equation for \( \delta \xi_i \) and insert the expression for \( \delta X_i \). We obtain

\[
\delta \xi_i = \frac{\partial f_1}{\partial q} (\delta X_i \cdot q_0) + \frac{\partial f_1}{\partial \mu} (\delta M_i) + \text{h.o.t.}
\]

\[
= \frac{\partial f_1}{\partial q} \left( h \sum_j a_{ij} \delta \xi_j \cdot q_0 \right) + \frac{\partial f_1}{\partial \mu} (\delta M_i) + \mathcal{O}(h^{r+3}) + \text{h.o.t.}
\]

\[
= \frac{\partial f_1}{\partial \mu} (\delta M_i) + \mathcal{O}(h^{r+3}) + \text{h.o.t.},
\]

and

\[
\delta X_i = h \sum_j a_{ij} \frac{\partial f_1}{\partial \mu} (\delta M_i) + \mathcal{O}(h^{r+3}) + \text{h.o.t.}
\]
Similarly, successively we get

\[
\delta n_i = \frac{\partial f_2}{\partial q} (\delta X_i \cdot q_0) + \frac{\partial f_2}{\partial \mu} (\delta M_i) + \text{h.o.t.}
\]

\[
= \frac{\partial f_2}{\partial \mu} (\delta M_i) + \mathcal{O}(h^{r+3}) + \text{h.o.t.},
\]

\[
\delta Y = h \sum_i b_i \delta \xi_i + \mathcal{O}(h^{r+3}) + \text{h.o.t.}
\]

\[
= h \frac{\partial f_1}{\partial \mu} \left( \sum_i b_i \delta M_i \right) + \mathcal{O}(h^{r+3}) + \text{h.o.t.},
\]

\[
\delta \Lambda = -\frac{1}{2} \text{ad}_{\delta Y}^* \mu_0 + h \sum_i b_i \delta n_i + \mathcal{O}(h^{r+3}) + \text{h.o.t.},
\]

\[
\delta \lambda_i = h b_i \left( \frac{B_{r+1}}{(r+1)!} h^r c_i \left( \text{ad}_{\xi_0}^* \right)^{r+1} \mu_0 + \frac{1}{2} \text{ad}_{\delta \xi_i}^* \mu_0 - \delta n_i \right) + \mathcal{O}(h^{r+2}) + \text{h.o.t.},
\]

\[
\delta M_i = \frac{B_{r+1}}{(r+1)!} h^{r+1} \left( -e_i^{r+1} + \sum_j a_{ij} \frac{b_j c_j}{b_i} \left( \text{ad}_{\xi_0}^* \right)^{r+1} \mu_0 + \mathcal{O}(h^{r+2}) \right).
\]

From the last equation, we see that

\[
\sum_i b_i \delta M_i = \mathcal{O}(h^{r+2}),
\]

which yields, successively,

\[
\delta Y = \mathcal{O}(h^{r+3}),
\]

\[
\sum_i b_i \delta n_i = \mathcal{O}(h^{r+2}),
\]

\[
\delta \Lambda = \mathcal{O}(h^{r+3}),
\]

concluding the proof. \(\square\)

An immediate consequence of the proof is that there exist methods in this class of arbitrarily high order. For instance, the Gauss methods [4, Section II.1.3] form a class of Runge–Kutta methods which achieve arbitrarily high order. Since these methods themselves are symplectic, \(\hat{a}_{ij} = b_j - b_j a_{ji} / b_i = a_{ij}\), and the variational method based on a Gauss method is a partitioned Runge–Kutta method with the same coefficients for both position and momentum. The variational method is equivalent to the Gauss method itself applied to the Hamiltonian ODE and has therefore the same order as the Gauss method itself. When \(r\) is large enough, the VRKMK method based on the coefficients of a Gauss method achieves the same order as the Gauss method.
5.2 Order of VCG Integrators

In this section, we will prove that there exist VCG integrators of any order. To show this, we will need the following lemma.

**Lemma 5.4 (Composition of VCG integrators)** Let \((A^{(1)}, b^{(1)})\), and \((A^{(2)}, b^{(2)})\) be the Butcher tableaux of Runge–Kutta methods with \(s^{(1)}\) and \(s^{(2)}\) stages, and \(\gamma\) a real number. The composition method formed by first applying the VCG method based on \((A^{(1)}, b^{(1)})\) with step-length \(\gamma h\) and then the VCG method based on \((A^{(2)}, b^{(2)})\) with step-length \((1 - \gamma)h\), is a VCG method with Butcher tableau

\[
\begin{align*}
&\gamma A^{(1)} & &0 \\
&\gamma b^{(1)}_1 & \cdots & \gamma b^{(1)}_{s^{(1)}} & (1 - \gamma)A^{(2)} \\
&\vdots & & \vdots \\
&\gamma b^{(1)}_1 & \cdots & \gamma b^{(1)}_{s^{(1)}} \\
&\gamma b^{(1)}_1 & \cdots & \gamma b^{(1)}_{s^{(1)}} & (1 - \gamma)b^{(2)}_1 & \cdots & (1 - \gamma)b^{(2)}_{s^{(2)}}
\end{align*}
\]

**Proof** Consider the discrete Lagrangians corresponding to the two VCG integrators that are to be composed,

\[
L_h^{(1)}(q_0, q_1) = h \sum_{i=1}^{s^{(1)}} b^{(1)}_i \ell \left( Q_i^{(1)}, \xi_i^{(1)} \right),
\]

with constraints

\[
Q_i^{(1)} = \exp \left( ha_{i1}^{(1)} \xi_i^{(1)} \right) \cdots \exp \left( ha_{is^{(1)}}^{(1)} \xi_i^{(1)} \right) q_0,
\]

\[
q_1 = \exp \left( hb_1^{(1)} \xi_1^{(1)} \right) \cdots \exp \left( hb_{s^{(1)}}^{(1)} \xi_1^{(1)} \right) q_0,
\]

and

\[
L_h^{(2)}(q_0, q_1) = h \sum_{i=1}^{s^{(2)}} b^{(2)}_i \ell \left( Q_i^{(2)}, \xi_i^{(2)} \right),
\]
with constraints
\[
Q_i^{(2)}(2) = \exp \left( h a_i^{(2)} \xi_i^{(2)} \right) \cdots \exp \left( h a_1^{(2)} \xi_1^{(2)} \right) q_0, \\
q_1 = \exp \left( h b_i^{(2)} \xi_i^{(2)} \right) \cdots \exp \left( h b_1^{(2)} \xi_1^{(2)} \right) q_0,
\]
as well as the composition discrete Lagrangian
\[
L_h^{(c)}(q_0, q_1) = L_\gamma^{(1)}(q_0, \bar{q}) + L_1^{(2)}(q_0, q_1),
\]
where \( \bar{q} \) is chosen so that \( L_h^{(c)} \) is extremized. It was proved in [5, Theorem 2.5.1] that the integrator corresponding to \( L_h^{(c)} \) is the composition method that results from composing the integrator corresponding to \( L_\gamma^{(1)} \) with the integrator corresponding to \( L_1^{(2)} \). Denote by \((A^{(c)}, b^{(c)})\) the Butcher tableau with \( s^{(c)} = s^{(1)} + s^{(2)} \) stages given above. Then,
\[
L_h^{(c)}(q_0, q_1) = h \gamma \sum_{i=1}^{s^{(1)}} b_i^{(1)} \ell(Q_i^{(1)}, \xi_i^{(1)}) + (1 - \gamma) \sum_{i=1}^{s^{(2)}} b_i^{(2)} \ell(Q_i^{(2)}, \xi_i^{(2)})
\]
\[
= h \sum_{i=1}^{s^{(c)}} b_i^{(c)} \ell(Q_i^{(c)}, \xi_i^{(c)})
\]
with constraints
\[
Q_i^{(c)} = \exp \left( h a_i^{(c)} \xi_i^{(c)} \right) \cdots \exp \left( h a_1^{(c)} \xi_1^{(c)} \right) q_0, \\
q_1 = \exp \left( h b_i^{(c)} \xi_i^{(c)} \right) \cdots \exp \left( h b_1^{(c)} \xi_1^{(c)} \right) q_0.
\]

**Proposition 5.5** There exist methods of any order among the VCG integrators.

*Proof* From [4, Section II.4], we know that if we compose a one-step method with itself using different step sizes, we can obtain arbitrarily high order, provided we choose the number of steps and the step sizes appropriately. Thus, we obtain VCG methods of any order by composition.

**Proposition 5.6** For VCG integrators applied to regular Lagrangian problems, the order conditions for first and second order are the same as for the underlying Runge–Kutta method, i.e.,
\[
\sum_{i=1}^s b_i = 1, \quad \text{and} \quad \sum_{i=1}^s b_i c_i = \frac{1}{2},
\]
where \( c_i = \sum_j a_{ij} \).
Proof We use variational order analysis, as presented in [5, Section 2.3]. Let the exact discrete Lagrangian be denoted

\[ L_h^E (q_0, q(h)) = \int_0^h L(q(t), \dot{q}(t)) \, dt, \quad \text{where} \quad q_0 = q(0). \]

The exact discrete Lagrangian can be expanded in powers of \( h \):

\[ L_h^E (q_0, q(h)) = \sum_{k=0}^\infty \frac{h^k}{k!} \left( \frac{d^k}{dh^k} L_h^E (q_0, q(h)) \bigg|_{h=0} \right). \]

From the right-trivialized HP equations (2) and (4), it is straightforward to show that

\[ \frac{d}{dt} L(q(t), \dot{q}(t)) = \frac{d}{dt} (\mu, \xi) = \langle \dot{\mu}, \dot{\xi} \rangle + \langle \mu, \ddot{\xi} \rangle = (f_2(z), f_1(z)) + \left\langle \mu, \frac{d}{dt} f_1(z) \right\rangle, \]

where \( z = (q, \mu) \) and \( f(z) = (f_1(z), f_2(z)) \). Thus, letting \( (\xi_0, n_0) = f(z_0) = f(q_0, \mu_0) \), and using \( \ell(q, \dot{q}) = L(q, \dot{q}) \), we get

\[ L_h^E (q_0, q(h)) = h\ell(q_0, \xi_0) + \frac{h^2}{2} \left( n_0, \xi_0 + \mu_0, \frac{\partial f_1}{\partial q} (\xi_0 \cdot q_0) + \frac{\partial f_1}{\partial \mu} (n_0 - \text{ad}_{\xi_0}^* \mu_0) \right) + O(h^3). \]

Similarly, we can expand the discrete Lagrangian in powers of \( h \) by using (18) together with \( Q_i \big|_{h=0} = q_0 \) and \( \xi_i \big|_{h=0} = \xi_0 \):

\[ L_h (q_0, q(h)) = \sum_{k=0}^\infty \frac{h^k}{k!} \left( \frac{d^k}{dh^k} L_h (q_0, q(h)) \bigg|_{h=0} \right) \]

\[ = \sum_{k=0}^\infty \frac{h^k}{k!} \left( \frac{d^k}{dh^k} \sum_{i=1}^s b_i \ell(Q_i, \xi_i) \bigg|_{h=0} \right) \]

\[ = h \left( \sum_i b_i \right) \ell(q_0, \xi_0) + \frac{h^2}{2} \left( 2 \sum_i b_i \frac{d}{dh} \ell(Q_i, \xi_i) \bigg|_{h=0} \right) + O(h^3). \]

By comparing equal powers of the two expansions, we see that the first-order condition is \( \sum_i b_i = 1 \), as in RK methods. The second term needs more work. We apply (4) together with \( n_i \big|_{h=0} = n_0 \) and \( M_i \big|_{h=0} = \mu_0 \) and get

\[ 2 \sum_i b_i \frac{d}{dh} \ell(Q_i, \xi_i) \bigg|_{h=0} = 2 \sum_i b_i \left( \left\langle n_i \cdot Q_i, \frac{dQ_i}{dh} \right\rangle + \left\langle M_i, \frac{d\xi_i}{dh} \right\rangle \right) \bigg|_{h=0} \]

\[ = 2 \sum_i b_i \left( n_0 \cdot q_0, \frac{dQ_i}{dh} \bigg|_{h=0} \right) + 2 \mu_0 \sum_i b_i \left( \frac{d\xi_i}{dh} \bigg|_{h=0} \right). \]
We calculate the derivatives of $Q_i$ and $\xi_i$ with respect to $h$ using (19):

$$\frac{dQ_i}{dh}\bigg|_{h=0} = \sum_j a_{ij} \xi_0 \cdot q_0 = c_i \xi_0 \cdot q_0,$$

$$\sum_i b_i \frac{d\xi_i}{dh}\bigg|_{h=0} = \sum_i b_i \left( \frac{\partial f_1}{\partial q} \circ \frac{dQ_i}{dh}\bigg|_{h=0} + \frac{\partial f_1}{\partial \mu} \circ \frac{dM_i}{dh}\bigg|_{h=0} \right).$$

We also need the derivative of $M_i$. In this expression, we apply (20) and simplify using the first-order condition:

$$\sum_i b_i \frac{dM_i}{dh}\bigg|_{h=0} = \left( 1 - \sum_i b_i c_i \right) n_0 - \frac{1}{2} \text{ad}_{\xi_0}^* \mu_0.$$

Putting these equations together, we obtain

$$2 \sum_i b_i \frac{d}{dh} \ell(Q_i, \xi_i)\bigg|_{h=0} = 2 \sum_i b_i c_i \langle n_0, \xi_0 \rangle$$

$$+ \left( \mu_0, 2 \sum_i b_i c_i \left( \frac{\partial f_1}{\partial q} (\xi_0 \cdot q_0) \right) \right. \left. + \frac{\partial f_1}{\partial \mu} \left( 2 \left( 1 - \sum_i b_i c_i \right) n_0 - \text{ad}_{\xi_0}^* \mu_0 \right) \right).$$

Thus, to get second order, we need the second-order RK condition $\sum_i b_i c_i = 1/2$. □

The computation for third order is similar, but much more complicated. We give the third-order conditions here, without proof.

**Proposition 5.7** For VCG methods applied to regular Lagrangian problems, using $b_i \hat{a}_{ij} + b_j a_{ji} = b_j b_j$ and $\hat{c}_i = \sum_{j=1}^s \hat{a}_{ij}$, the conditions for third order are

$$\sum_{i=1}^s b_i c_i^2 = \frac{1}{3},$$

$$\sum_{i=1}^s \sum_{j=1}^s b_i a_{ij} c_j = \frac{1}{6},$$

$$\sum_{i=1}^s b_i c_i \left( \sum_{j=1}^{i-1} b_j + \frac{b_i}{2} \right) = \frac{1}{3},$$

$$\sum_{i=1}^s b_i \hat{c}_i^2 = \frac{1}{3}.$$
\[
\sum_{i=1}^{s} b_i \hat{c}_i \left( \sum_{j=1}^{i-1} b_j + \frac{b_i}{2} \right) = \frac{1}{3},
\]

\[
\sum_{i=1}^{s} b_i^3 = 0.
\]

The first two conditions come from standard RK methods, the third condition comes from CG methods, and the fourth comes from SPRK methods, while the final two conditions are new. The last condition also appears in the order conditions for compositions of one-stage RK methods. It is noteworthy that the final condition forces at least one of the weights \(b_i\) to be negative.

The general order theory for VCG integrators is not complete and needs further study.

### 6 Numerical Tests

To test our methods, we constructed a Hamiltonian test problem which we call “dipole on a stick” (see Fig. 1). The problem models a pendulum consisting of a long straight rod of length 1, with one end fixed (but freely rotating) at the origin, and a shorter rod of length \(2\alpha\) with its center attached perpendicularly to the long rod at the other end. At each of the end points of the shorter rod, there are charged particles with masses \(m/2\) and electric charges \(\pm q\). The rods are assumed to be massless. The pendulum is affected by gravity in the negative \(e_3\)-direction and the electric field generated by a charged particle at position \(z = (0, 0, -3/2)^T\) of charge \(\beta\). The physical constants for specific gravity and electric force are set equal to 1. We chose this test problem, so that it would have chaotic behavior and conserved energy, with \(\text{SO}(3)\) as configuration space.

If we let \(y_+(t), y_-(t)\) denote the positions of the positive and negative charge, respectively, the position of the pendulum can be described uniquely by the matrix \(g(t) \in \text{SO}(3)\) such that \(y_\pm(t) = g(t)y_\pm^0\), where \(y_\pm^0 = (0, \pm\alpha, -1)^T\) is the position of the two charged particles in reference or body coordinates. Using the standard identification of \(\mathfrak{so}(3)\) with \(\mathbb{R}^3\) and of \(\mathfrak{so}(3)^*\) with \(\mathbb{R}^3\) via the standard inner product,

![Fig. 1 Dipole on a stick](image)
the state of the system \((g, \mu)\) can be represented with \(g \in SO(3)\) as a \(3 \times 3\) real matrix and \(\mu \in so(3)^*\) as a vector in \(\mathbb{R}^3\).

The right-trivialized Hamiltonian of this system is

\[
\mathcal{H}(g, \mu) = \frac{1}{2} \mu^T g I^{-1} g^T \mu + me_3^T g e_3 + q \beta \left( \|gy^0_- - z\|^{-1} - \|gy^0_+ - z\|^{-1} \right),
\]

where \(I = m \text{diag}(1 + \alpha^2, 1, \alpha^2)\) is the inertia tensor of the pendulum.

6.1 Order Tests

The VRKMK methods that were tested are based on the 1-, 2-, and 3-stage Gauss methods and Kutta’s third-order method. These methods are defined by the Butcher tableaux and cutoff parameters in Table 1. These methods can be shown to satisfy the extra order conditions for variational integrators to their respective orders.

The order of the VCG methods was also tested. To obtain higher order, symmetric composition of the midpoint method (which is symmetric, see Example 3.3) as described in [4, Section V.3.2] was used. The Butcher tableaux of the resulting methods are the same as those of the fourth- and sixth-order diagonally implicit Runge–Kutta methods (DIRK) shown in Table 2. The parameters \(\gamma_1, \ldots, \gamma_4\) were derived by Yoshida [21].

The methods were implemented in MATLAB, using a modified version of the DiffMan package [22] for defining Lie algebra and Lie group classes and functions on these spaces. The sets of nonlinear equations (17) and (19) were solved by fixed-point iteration. The iteration was terminated when the norm of the residual became less than \(10^{-11}\).

In these tests, we have used the data

\[
m = q = \beta = 1, \\
\alpha = 0.1, \\
g(0) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \\
\mu(0) = g(0) I g(0)^T e_2.
\]

The initial data \(\mu(0)\) are chosen so that the first component of \(f(g(0), \mu(0))\) is \(e_2\).

The errors in \(\mu(0.5)\) and \(g(0.5)\) with respect to a reference solution are shown in Fig. 2. The errors plotted are \(\|\mu - \mu_{\text{ref}}\|_2 + \|g - g_{\text{ref}}\|_2\), where the first \(\|\cdot\|_2\) is the Euclidean vector norm and the second is the subordinate matrix norm. The reference solution was calculated using the sixth-order VRKMK method with step size \(h = 10^{-3}\). The dashed lines are reference lines for the appropriate orders and are the same lines in the two plots. As is evident from the plots, errors from fixed-point iteration dominate the errors for the sixth-order methods when \(h\) is small, and the methods appear to obtain their theoretical order.
Table 1 Butcher tableaux of the RKMK methods tested

|       | 0   | 0   | 0   | 0   | 1   | 1   | 2   | 0   |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|
| $r = 0$ | 0.5 | 0.5 |     |     | 1   |     |     |     |

(a) Second order Gauss method

(b) Kutta’s third order method

|       | $\frac{1}{2} - \frac{\sqrt{3}}{6}$ | $\frac{1}{4}$ | $\frac{1}{4} - \frac{\sqrt{3}}{6}$ | $\frac{1}{2} + \frac{\sqrt{3}}{6}$ | $\frac{1}{4} + \frac{\sqrt{3}}{6}$ | $\frac{1}{4}$ |
|-------|------------------------------------|---------------|------------------------------------|------------------------------------|----------------|------------|
| $r = 2$ | $\frac{1}{2} - \frac{\sqrt{15}}{10}$ | $\frac{5}{36}$ | $\frac{2}{9} - \frac{\sqrt{15}}{15}$ | $\frac{5}{36}$ | $\frac{2}{9} + \frac{\sqrt{15}}{15}$ | $\frac{5}{36}$ |

(c) Fourth order Gauss method

(d) Sixth order Gauss method
Table 2  Butcher tableaux of the VCG methods tested

|       | \( \frac{1}{2} \gamma_1 \) | \( \frac{1}{2} \gamma_1 \) | 0 | 0 | 0 | 0 |
|-------|-----------------------------|-----------------------------|---|---|---|---|
| \( \frac{1}{2} \gamma_1 \) | \( \gamma_1 + \frac{1}{2} \gamma_2 \) | \( \gamma_1 \) | \( \frac{1}{2} \gamma_2 \) | 0 | 0 | 0 | 0 |
| \( \gamma_1 + \gamma_2 + \frac{1}{2} \gamma_3 \) | \( \gamma_1 \) | \( \gamma_2 \) | \( \frac{1}{2} \gamma_3 \) | 0 | 0 | 0 | 0 |
| \( \frac{1}{2} \gamma_1 \) | \( 1 - \gamma_1 \) | \( \gamma_1 \gamma_2 \gamma_3 \gamma_4 \) | \( \frac{1}{2} \gamma_4 \) | 0 | 0 | 0 | 0 |

\( \gamma_1 \gamma_2 \gamma_3 \gamma_4 \gamma_3 \gamma_2 \frac{1}{2} \gamma_1 \)

\( \gamma_1 = \frac{1}{2-2^{1/3}} \), \( \gamma_2 = \frac{-2^{1/3}}{2-2^{1/3}} \)

(a) Second order midpoint method

(b) Fourth order DIRK method based on triple jump

\( \gamma_1 = 0.78451361047755726381949763 \), \( \gamma_2 = 0.23557321335935813368479318 \)
\( \gamma_3 = -1.17767998417887100694641568 \), \( \gamma_4 = 1.31518632068391121888424973 \)

(c) Sixth order DIRK method
Analytically, the second-order VRKMK and VCG methods are actually identical. The implementations of the two methods are different, as the nonlinear equations are set up in slightly different manners. The result of this is that the numerical solutions differ slightly. For this numerical test, the error constants of the VRKMK methods are smaller than those of the VCG methods.

6.2 Long-Term Behavior

The long-term behavior of the methods was also investigated. In Fig. 3, the energy error of the numerical solution is plotted over the time span \((0, 1000)\). We have used step size \(h = 0.01\) (10^5 integration steps). Only the second- and fourth-order methods were tested on this time span. As can be seen from the plots, the energy error is small, approximately \(10^{-3}\) for both second-order methods, and approximately \(10^{-7}\) for the fourth-order VRKMK method and about \(10^{-5}\) for the fourth-order VCG method.

7 Future Work

The reformulation of RKMK methods with a Padé approximation of \(d\exp^{-1}\) is briefly discussed in Sect. 4.1.1. This reformulation makes it possible to eliminate the Lagrange multipliers \(\lambda_j\), which is beneficial for computational efficiency. We expect that the proof of the order of VRKMK methods, Theorem 5.3, will carry over to these methods without complications. Implementation and study of this approach would make the variational RKMK methods more competitive in terms of computational cost.

Another class of Lie group methods is formed by the commutator-free Lie group methods. The approach described in this article can easily be used to formulate symplectic Lie group methods based on commutator-free methods. Formulation, implementation, and study of variational commutator-free methods are aspects that can be pursued in the future.

A desirable result would be generalization of these integrators to homogeneous spaces. This has proven to be more difficult than one could hope. In general, the
problem arises due to isotropy. If $M$ is a homogeneous $G$-space with $\dim(M) < \dim(G)$, then the infinitesimal action at a point $z \in M$,

$$g \ni \xi \mapsto \frac{\partial}{\partial t} \exp(t\xi) \cdot z \in T_z M,$$
is not injective. Therefore, to identify a vector in $T_z M$ with some element in $g$, a choice has to be made.

The main idea of variational integration is to minimize the action. Inspired by this, one could attempt the following approach, sketched out for a variational method based on the one-stage $\theta$-method for $0 \leq \theta \leq 1$. Assume the action is from the left, and denote the action as $g \cdot q$, and the infinitesimal action as $\xi \cdot q$. Let $\ell(q, \xi) = L(q, \xi \cdot q)$ be the “trivialized” Lagrangian and use the discrete Lagrangian

$$L_h(q_0, q_1) = \min_{\xi} h \ell(\exp(h\xi) \cdot q_0, \xi)$$

where the minimum is taken over all $\xi$ such that $\exp(h\xi) \cdot q_0 = q_1$. If the minimizing equation can be solved, this discrete Lagrangian can be used to construct symplectic integrators. However, the following example shows that in some cases, the minimizing equation has no solution. Let $M = \mathbb{R}$ and the group action that of affine functions $\mathbb{R} \to \mathbb{R}$, i.e., for $(a, b) \in (\mathbb{R} \setminus \{0\}) \times \mathbb{R} = G$, $(a, b) \cdot q = aq + b$. Let the Lagrangian be that of a free particle, $L(q, \dot{q}) = \frac{1}{2} \dot{q}^2$. In this case, it turns out that the minimizing problem (27) can be expressed as an unconstrained one-dimensional problem,

$$L_h(q_0, q_1) = \min_{x \in \mathbb{R}} h \left( \frac{x e^{hx}}{e^{hx} - 1} \right)^2 (q_1 - q_0)^2.$$ 

However, this has no solution if $q_1 \neq q_0$, so $L_h(q_0, q_1)$ is not defined. Furthermore, for $0 < \theta < 1$, the expression has a maximizer, so a naive solution to the extremization problem would return the maximizing solution. The symplectic method based on such a solution is not even consistent.

8 Conclusion

In this article, a set of equations defining symplectic integrators for ODEs on $T^*G$ were presented, as well as two classes of integrators using these equations. The integrators obtained are formulated intrinsically on $T^*G$, and any drift away from the manifold in numerical solutions is due to round-off errors. The integrators were developed as variational methods for Lagrangian problems and are therefore symplectic when applied to Hamiltonian differential equations. Both classes that were studied were shown to contain methods of arbitrarily high order, although the computational cost per time step increases with the order. Effective implementation of the methods has not been a major goal in this article, we have instead focused on the properties of these methods.

The two classes of symplectic methods are based on, respectively, the Runge–Kutta–Munthe-Kaas methods and the Crouch–Grossman methods. The methods have a partitioned structure where the position on the Lie group is integrated by the Lie group method, while the momentum is integrated by formulae which involve various functions on $g^*$. We can therefore say that these methods are partitioned Lie group methods and are Lie group methods in a wide understanding of that term. To the
knowledge of the authors, this is the first time that symplectic partitioned Lie group methods have been presented and studied in the level of detail done in this article.

Acknowledgments We would like to thank our supervisor, Brynjulf Owren, for encouragement and many helpful discussions. We would also like to thank Klas Modin for fruitful discussions leading to the writing of this paper. Finally, we would like to thank the two anonymous referees for helpful comments and suggestions. The research was supported by the Research Council of Norway and by a Marie Curie International Research Staff Exchange Scheme Fellowship within the 7th European Community Framework Programme.

References

1. Peter E. Crouch and Robert L. Grossman. ‘Numerical integration of ordinary differential equations on manifolds’. In: J. Nonlinear Sci. 3.1 (1993), pp. 1–33, doi: 10.1007/BF02429858.
2. Hans Z. Munthe-Kaas. ‘High order Runge–Kutta methods on manifolds’. In: Proceedings of the NSF/CBMS Regional Conference on Numerical Analysis of Hamiltonian Differential Equations (Golden, CO, 1997), Vol. 29. 1. 1999, pp. 115–127, doi: 10.1006/S0168-9274(98)00030-0.
3. Arieh Iserles, Hans Z. Munthe-Kaas, Syvert Paul Nørsett and Antonella Zanna. ‘Lie-group methods’. In: Acta numerica, 2000. Vol. 9. Acta Numer. Cambridge: Cambridge Univ. Press, 2000, pp. 215–365. doi:10.1017/S0962492900002154.
4. Ernst Hairer, Christian Lubich and Gerhard Wanner. Geometric numerical integration. Structure-preserving algorithms for ordinary differential equations. 2nd ed. Vol. 31. Springer Series in Computational Mathematics. Berlin: Springer-Verlag, 2006, pp. xviii+644. doi: 10.1007/3-540-30666-8.
5. Jerrold Eldon Marsden and Matthew West. ‘Discrete mechanics and variational integrators’. In: Acta Numer. 10 (2001), pp. 357–514. doi:10.1017/S096249290100006X.
6. Melvin Leok. Variational Integrators. In: Encyclopedia of Applied and Computational Mathematics. Ed. by Björn Engquist. Springer-Verlag, 2013.
7. Nawaf M. Bou-Rabee and Jerrold Eldon Marsden. ‘Hamilton–Pontryagin integrators on Lie groups. I. Introduction and structure-preserving properties’. In: Found. Comput. Math. 9.2 (2009), pp. 197–219. doi:10.1007/s10208-008-9030-4.
8. Elena Celledoni, Håkon Marthinsen and Brynjulf Owren. ‘An introduction to Lie group integrators – basics, new developments and applications’. In: J. Comput. Phys. 257 (2014), pp. 1040–1061. doi:10.1016/j.jcp.2012.12.031.
9. Melvin Leok. ‘Foundations of computational geometric mechanics’. PhD thesis. California Institute of Technology, 2004. URL: http://resolver.caltech.edu/CaltechETD:etd-03022004-000251.
10. James Hall and Melvin Leok. Lie Group Spectral Variational Integrators. 2014. arXiv: 1402.3327 [math.NA]. Submitted.
11. Vladimir Igorevich Arnold and Boris A. Khesin. Topological methods in hydrodynamics. Vol. 125. Applied Mathematical Sciences. Springer-Verlag, New York, 1998, pp. xvi+374. doi:10.1007/b97593.
12. Kenth Engø. ‘Partitioned Runge–Kutta methods in Lie–group setting’. In: BIT 43.1 (2003), pp. 21–39. doi:10.1023/A:1023668015087.
13. Yuri Borisovich Suris. ‘Hamiltonian methods of Runge–Kutta type and their variational interpretation’. In: Mat. Model. 2.4 (1990), pp. 78–87. URL: http://mi.mathnet.ru/eng/mmm2356.
14. Melvin Leok and Jingjing Zhang. ‘Discrete Hamiltonian variational integrators’. In: IMA J. Numer. Anal. 31.4 (2011), pp. 1497–1532. doi:10.1093/imanum/drq027.
15. Elena Celledoni, Arne Marthinsen and Brynjulf Owren. ‘Commutator-free Lie group methods’. In: Future Generation Computer Systems 19.3 (2003), pp. 341–352. doi:10.1016/S0167-739X(02)00161-9.
16. Brynjulf Owren. ‘Order conditions for commutator-free Lie group methods’. In: J. Phys. A 39.19 (2006), pp. 5585–5599. doi:10.1088/0305-4470.
17. Jerrold Eldon Marsden and Tudor S. Ratiu. Introduction to mechanics and symmetry. A basic exposition of classical mechanical systems. 2nd ed. Vol. 17. Texts in Applied Mathematics. Springer-Verlag, New York, 1999, pp. xviii+582. doi: 10.1007/978-0-387-21972-5.
18. Brynjulf Owren and Arne Marthinsen. ‘Runge–Kutta methods adapted to manifolds and based on rigid frames’. In: BIT 39.1 (1999), pp. 116–142. doi:10.1007/A:102235426017.
19. George W. Patrick and Charles Cuell. ‘Error analysis of variational integrators of unconstrained
Lagrangian systems’. In: *Numer. Math.* 113.2 (2009), pp. 243–264. doi:10.1007/s00211-009-0245-3.
20. Ander Murua. ‘On order conditions for partitioned symplectic methods’. In: *SIAM J. Numer. Anal.*
34.6 (1997), pp. 2204–2211. doi:10.1137/S0036142995285162.
21. Haruo Yoshida. ‘Construction of higher order symplectic integrators’. In: *Phys. Lett. A* 150.5–7 (1990),
pp. 262–268. doi:10.1016/0375-9601(90)90092-3.
22. Kenth Engø, Arne Marthinsen and Hans Z. Munthe-Kaas. ‘DiffMan: an object-oriented MATLAB
toolbox for solving differential equations on manifolds’. In: *Appl. Numer. Math.* 39.3–4 (2001). Special
issue: Themes in geometric integration, pp. 323–347. doi:10.1016/S0168-9274(00)00042-8.