VARIATIONAL DEVELOPMENT OF THE SEMI-SYMPLECTIC GEOMETRY OF NONHOLONOMIC MECHANICS

GEORGE W. PATRICK[†]

Applied Mathematics and Mathematical Physics,
Department of Mathematics and Statistics, University of Saskatchewan,
Saskatoon, Saskatchewan, S7N5E6, Canada
(e-mail: patrick@math.usask.ca)

(Received August 21, 2006)

The geometry of constrained Lagrangian systems is developed using the Lagrange–d'Alembert principle, extending the variational approach of Marsden, Patrick and Shkoller [26] from holonomic to nonholonomic systems. It emerges that the instrinsic geometry of nonholonomic systems corresponds to the geometry of the distributional Hamiltonian systems of Śniatycki [35], here called semi-Hamiltonian. The principle physical reason that nonholonomic systems exhibit non-symplectic dynamics is exposed, leading to curvature conditions for the presence of holonomic subsystems. The underlying geometry of semi-Hamiltonian systems is semi-symplectic. An abstract exposition of the semi-symplectic category is developed. This is closed under a reduction scheme able to incorporate conserved quantities which are not momenta but are often structurally implied by symmetry. The variational development is continued to include nonlinear constraints irrespective of whether or not they are obtained from Chetaev's rule. Even though these systems are not semi-Hamiltonian, their geometry is still semi-symplectic.

Keywords: nonholonomic mechanics, distributional Hamiltonian system, semi-symplectic, variational methods.

1. Introduction

Holonomic mechanics shines in the company of abstract symplectic geometry of phase space, in a way which naturally diminishes impositions from tangent and cotangent fibrations. Analogously, nonholonomic mechanics shines in the company of semi-symplectic geometry.

It is best appreciated by starting from the variational principle. A manifold Q and a function $L : TQ \to \mathbb{R}$ define a (holonomic) *Lagrangian system*. The *action functional* assigns to curves $q(t) \in Q$ the number

$$S\big(q(t)\big) \equiv \int_a^b L\big(q'(t)\big)\, dt.$$

Hamilton's principle states that the time evolution is given by the curves $q(t)$ such

[†]Supported in part by the Natural Sciences and Engineering Research Council, Canada.

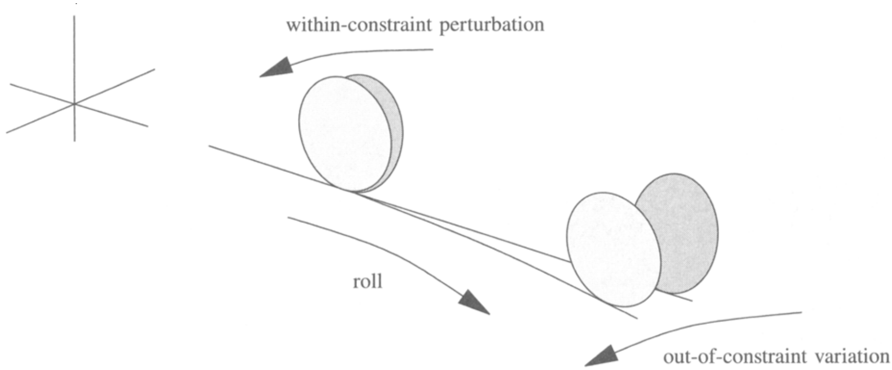

Fig. 1. Tipping a rolling disk causes it to veer through configurations inconsistent with the rolling constraint.

that

$$\boldsymbol{d}S\big(q(t)\big)\cdot\delta q(t)\equiv\frac{d}{d\epsilon}\bigg|_{\epsilon=0}S\big(q_\epsilon(t)\big)=0,$$

where

$$\delta q\equiv\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}q_\epsilon(t)$$

for all smooth curves $q_\epsilon(t)$ satisfying $q_0(t)=q(t)$ and the fixed-endpoint condition $q_\epsilon(a)=q(a), q_\epsilon(b)=q(b)$. In coordinates, the usual integration by parts gives

$$\boldsymbol{d}S\big(q(t)\big)\cdot\delta q(t)=\frac{\partial L}{\partial\dot q^i}\delta q^i\bigg|_a^b+\int_a^b\left(\frac{\partial L}{\partial q^i}-\frac{d}{dt}\frac{\partial L}{\partial\dot q^i}\right)\delta q^i\,dt.$$

The first, boundary term $(\partial L/\partial\dot q^i)\delta q^i$ is an evaluation of the Lagrange one-form $\theta_L=(\partial L/\partial\dot q^i)dq^i$. This vanishes under the fixed-endpoint condition, but it can be retained by allowing the endpoints in q_ϵ to vary. It can then be isolated by restricting the action functional to the solutions of the Euler–Lagrange equations, because under that restriction the second, non-boundary term vanishes. Isolation of the boundary term isolates the Lagrange one-form—a foundational differential-geometric object. Marsden, Patrick and Shkoller [26] use this development to clarify the structures of Moser–Veselov discrete mechanics, to identify the analogue of flow-symplecticity for field theories, and to create multisymplectic discretizations for field theories.

In this work, an analogous variational development is applied to *constrained Lagrangian systems*, in which the time evolution curves $q(t)$ are determined by the *Lagrange–d'Alembert principle*

$$\boldsymbol{d}S\big(q(t)\big)\cdot\delta q(t)=0\tag{1}$$

for all

$$\delta q(t)\in\mathcal{D},\qquad q'(t)\in\mathcal{D},$$

where $\mathcal{D}$ is a distribution on $\boldsymbol{T}Q$. If $\mathcal{D}$ is involutive (*holonomic system*) then (1) corresponds to a set of symplectic Hamiltonian systems, one for each leaf of

$\mathcal{D}$. If $\mathcal{D}$ fails to be involutive (*nonholonomic system*), then (1) is a *skew variational principle*, meaning that the objective S is differentiated in directions other than tangent directions of the constraint $q'(t) \in \mathcal{D}$.

Mainly, the development in this work provides two outstanding insights about nonholonomic systems:

1. Configuration-projected variations of nonholonomic systems are not generally within-constraint, even if they arise from within-constraint perturbations of initial conditions. For example, a disk that is rolling vertically cannot be directly rolled along configurations obtained at a fixed time after tipping the disk (Fig. 1). It turns out that this physical feature of the rolling-disk system obstructs the isolation of the boundary term in the variational development and fully accounts for the system's nonsymplectic geometry. *Nonholonomic systems exhibit nonsymplectic dynamics inasmuch as within-constraint perturbations yield out-of-constraint variations, because it is exactly this that obstructs the isolation of the boundary part of the Lagrange–d'Alembert principle.*
2. A *semi-Hamiltonian system* is a manifold $\mathcal{P}$, a distribution $\mathcal{K}$, a nondegenerate antisymmetric $\binom{2}{0}$ tensor ω on $\mathcal{K}$, and a function $H : \mathcal{P} \to \mathbb{R}$. The time evolution is given by the integral curves of the vector field Y_H in $\mathcal{K}$ defined by the semi-Hamilton equations
$$\boldsymbol{i}_{Y_H}\omega(v) = \boldsymbol{d}H(v) \qquad \text{for all} \quad v \in \mathcal{K}.$$
Semi-Hamiltonian systems have been defined in [35] where they are called *distributional Hamiltonian systems.* As is proved in [4], and as naturally emerges from the development, *constrained Lagrangian systems are semi-Hamiltonian systems.* The economy of the semi-symplectic view has been emphasized in the work of Śniatycki and this is reinforced here.

The development also provides concise derivations of the semi-symplectic equations for the evolution vector field and, in the presence of symmetry, of the nonholonomic momentum equation.

It is not necessary to presuppose a fiber bundle $\pi : \mathcal{Q} \to \mathcal{R}$ on which $\mathcal{D}$ is the horizontal space of an Ehresmann connection, as is emphasized for example in [6, 31]. An alternative to the semi-symplectic framework posits a manifold $\hat{\mathcal{P}} \supseteq \mathcal{P}$ and a distribution $\mathcal{F}$ on $\hat{\mathcal{P}}$. The semi-Hamilton equations are replaced by $Y_H \in \boldsymbol{T}P$ and $\boldsymbol{i}_{Y_H}\omega(v) = \boldsymbol{d}H(v)$ for all $v \in \mathcal{F}$. This framework is emphasized for example in [9, 10, 13]. It raises the ambient space in which the nonholonomic phase space typically lies to a definitional status, so taking a more explicitly constrained stance. The variational development in this work seems to diminish the ambient apparatus to a subordinate special case status rather than a definitional one. In [5] there is an exposition based on Dirac brackets.

Variations arise from the flow and they are in-constraint under certain Lie bracket conditions. Those are related to curvature, so the development reveals curvature conditions for the presence of holonomic subsystems. This seems to be an overlooked aspect. Holonomic subsystems actually occur—for example, nontranslating, spinning motions form a holonomic subsystem for the vertically rolling disk. Nonholonomic geometric integrators for nonholonomic systems ought to be shown to address symplectic subsystems, or they may not be preserving structures that are present. Holonomic subsystems of nonholonomic systems should not be unexpected and they might provide opportune ground to perturb from.

Noether's theorem provides a one-to-one correspondence between symmetries and momenta of symmetric Hamiltonian systems. The correspondence is broken in semi--Hamiltonian systems by the disparate notions of "preservation of $\mathcal{K}$", corresponding to "symmetry", and "membership in $\mathcal{K}$", corresponding to "conserved momentum". This complicates nonholonomic reduction and can lead to conserved quantities that do not generate symmetries and are functionally independent of conserved momenta. Bookkeeping for these *orphaned conserved quantities* can be arranged by including a subbundle $\mathcal{K}^\circ$ in the definition of a semi-Hamiltonian manifold, and restricting to Hamiltonians with differential in $\mathcal{K}^\circ$. The level sets of the orphaned conserved quantities become the reachability sets of semi-Hamiltonian vector fields of functions with differential in $\mathcal{K}^\circ$. Including $\mathcal{K}^\circ$ raises the semi-symplectic category to a central infrastructure supporting constrained Lagrangian systems because including $\mathcal{K}^\circ$ closes the semi-symplectic category under reduction in the regular case.

Distributional constraints such as $\mathcal{D}$ are called *linear*. In the generalization to *nonlinear constraints*, $\mathcal{D}$ is replaced with a submanifold $\mathcal{C} \subseteq \boldsymbol{T}\mathcal{Q}$, and the time-evolution curves $q(t)$ are determined by

$$
\begin{aligned}
&\boldsymbol{d}S\big(q(t)\big)\cdot\delta q(t)=0 \qquad \text{for all} \quad \delta q(t)\in\mathcal{W}_{q'(t)},\\
&q'(t)\in\mathcal{C},
\end{aligned}
\tag{2}
$$

where $\mathcal{W}$ is a subbundle of the pull-back bundle $(\tau_{\mathcal{Q}}|\mathcal{C})^*(\boldsymbol{T}\mathcal{Q})$. $\mathcal{W}$ smoothly associates to each $v_q \in \mathcal{C}$ a subspace $\mathcal{W}_{v_q}$ of $\boldsymbol{T}_q\mathcal{Q}$, the annihilator of which is the space of possible constraint forces at the dynamic state v_q. The *Chetaev rule* for example posits that $\mathcal{W}$ corresponds to the vertical subbundle of $\boldsymbol{T}\mathcal{C}$, but other choices might be indicated by the physical particulars. Nonlinearly constrained systems can be associated with external forcing, so they cannot in general be semi-Hamiltonian because they do not conserve energy, which every semi-Hamiltonian system does. Nevertheless, by a variational development (Section 6), the *geometry* of nonlinearly constrained systems is semi-symplectic because their evolution vector fields are naturally a semi-symplectic part plus a forcing part. If $\mathcal{W}$ contains the diagonal of $\mathcal{C}\times\mathcal{C}$ then the forcing part is zero, the system is semi-Hamiltonian, and therefore is energy conserving. This extends a well-known result that, within the Chetaev rule, energy is conserved if the Liouville vector field is tangent to the constraint $\mathcal{C}$. Fully nonlinear constraints, although sometimes specifically avoided in the literature, are not structurally different or more complicated than *affine constraints*, where $\mathcal{C}$ is a bundle with fibers which are affine subspaces of the fibers of $\boldsymbol{T}\mathcal{Q}$. Also, arbitrary $\mathcal{W}$ are not structurally different or more complicated then the Chetaev choice.

The theory of nonholonomic systems has an extensive literature; recent reviews include [7, 12]. The most significant debts owed by this work are:

- [4, 35], for the identification of semi-Hamiltonian systems and their reductions;
- [6], for the theory of symmetric systems, the nonholonomic momentum equation, and the emphasis of the Lagrange–d'Alembert principle; and
- [22, 23], for the differential-geometric definition of nonlinearly constrained nonholonomic systems.

Nonholonomic systems have been studied for centuries and there has been a cascade of results from the early 1990s. This work is integrative but not a review and only some of the relationships with the more recent literature are discussed.

The notation used here is all but identical to that of [1]. If $\mathcal{M}$ is a manifold, the vertical lift at $v_q \in \boldsymbol{T}_q\mathcal{M}$ of $w_q \in \boldsymbol{T}_q\mathcal{M}$ is denoted by

$$\mathrm{vert}_{v_q} w_q \equiv \left.\frac{d}{dt}\right|_{t=0} v_q + t w_q.$$

The canonical involution of $\boldsymbol{T}\boldsymbol{T}\mathcal{M}$ by $s_{\mathcal{M}}$, so that, for all smooth curves $c(s,t)$,

$$\frac{\partial}{\partial s}\frac{\partial}{\partial t} c(s,t) = s_{\mathcal{M}} \frac{\partial}{\partial t}\frac{\partial}{\partial s} c(s,t).$$

Where unambiguous, juxtaposition is used for binary operations. For example, the action of some group element g on some p may be abreviated gp, and the infinitesimal generator of some Lie algebra element ξ at p may be abbreviated ξp. If $\mathcal{E}$ is a bundle and p is in the base then the fiber over p is denoted $\mathcal{E}_p$.

2. Distributions, curvature, forms

It is efficient to begin by collecting some general results about distributions and the calculus of forms defined on them.

Let $\mathcal{M}$ be a manifold and $\mathcal{D}$ be a distribution on $\mathcal{M}$. The following lemma gives equivalent conditions for $\mathcal{D}$ to be involutive in terms of a distribution on $\boldsymbol{T}\mathcal{M}$ which is derived from $\mathcal{D}$.

LEMMA 2.1. *$\widehat{\mathcal{D}} \equiv (\boldsymbol{T}\tau_{\mathcal{M}})^{-1}\mathcal{D} \cap \boldsymbol{T}\mathcal{D}$ is a distribution on $\mathcal{D}$ which has fiber dimension twice the fiber dimension of $\mathcal{D}$. The following are equivalent:*

1. *$\mathcal{D}$ is involutive,*
2. *$\widehat{\mathcal{D}}$ is involutive,*
3. *$s_{\mathcal{M}}\widehat{\mathcal{D}} = \widehat{\mathcal{D}}$,*
4. *$\widehat{\mathcal{D}} \subseteq s_{\mathcal{M}}\boldsymbol{T}\mathcal{D}$,*

in which case the maximal integral submanifolds of $\widehat{\mathcal{D}}$ are exactly the tangent spaces of the maximal integral submanifolds of $\mathcal{D}$.

Proof: The lemma is local, so work in coordinates x^i, and let X_α be vector fields such that $\{X_\alpha(x)\}$ is a basis of the fiber of $\mathcal{D}$ at x. If $f^\alpha(t)$ are functions and $x(t)$ is a curve, then $f^\alpha(t)X^i_\alpha(x(t))$ is a general curve in $\mathcal{D}$, and

$$\frac{d}{dt} f^\alpha(t)X^i_\alpha\big(x(t)\big) = \frac{df^\alpha}{dt}(t)X^i_\alpha\big(x(t)\big) + f^\alpha(t)\,\frac{\partial X^i_\alpha}{\partial x^j}\big(x(t)\big)\,\frac{dx^j}{dt},$$

so that

$$T\mathcal{D} = \left\{ (x^i, f^\alpha X^i_\alpha, \dot{x}^i, \dot{v}^i) : \dot{v}^i = \dot{f}^\alpha X^i_\alpha + f^\alpha \dot{x}^j \frac{\partial X^i_\alpha}{\partial x^j};\ f^\alpha, \dot{f}^\alpha,\ \dot{x}^i \in \mathbb{R} \right\},$$

and putting $\dot{x}^i = g^\alpha X^i_\alpha$,

$$\widehat{\mathcal{D}} = \left\{ (x^i, f^\alpha X^i_\alpha, g^\alpha X^i_\alpha, \dot{v}^i) : \dot{v}^i = \dot{f}^\alpha X^i_\alpha + f^\alpha g^\beta X^j_\beta \frac{\partial X^i_\alpha}{\partial x^j};\ f^\alpha, g^\alpha, \dot{f}^\alpha \in \mathbb{R} \right\}.$$

So for fixed base point $(x^i, f^\alpha X^i_\alpha)$, the fiber of $\widehat{\mathcal{D}}$ has a basis

$$X^i_\alpha \frac{\partial}{\partial v^i}, \qquad X^i_\alpha \frac{\partial}{\partial x^i} + f^\beta X^j_\alpha \frac{\partial X^i_\beta}{\partial x^j} \frac{\partial}{\partial v^i}, \tag{3}$$

which number twice the dimension of the fiber of $\mathcal{D}$.

(1) $\Leftrightarrow$ (2): $\widehat{\mathcal{D}} \subset T\mathcal{D}$ so $\widehat{\mathcal{D}}$ is involutive if and only if it is involutive as a distribution on the manifold $\mathcal{D}$. In the coordinates (x^i, f^α) of $\mathcal{D}$, $\tau_{\mathcal{M}}$ is $(x^i, f^\alpha) \mapsto x^i$, and so $\widehat{\mathcal{D}} = \{(x^i, f^\alpha, g^\alpha X^i_\alpha, \dot{f}^\alpha)\}$ which is involutive if and only if $\mathcal{D}$ is involutive.

(1) $\Leftrightarrow$ (3): Since

$$s_{\mathcal{M}} \left\{ (x^i, g^\alpha X^i_\alpha, f^\alpha X^i_\alpha, \dot{v}^i) : \dot{v}^i = \dot{f}^\alpha X^i_\alpha + f^\alpha g^\beta X^j_\beta \frac{\partial X^i_\alpha}{\partial x^j},\ f^\alpha, g^\alpha, \dot{f}^\alpha \in \mathbb{R} \right\}$$
$$= \left\{ (x^i, f^\alpha X^i_\alpha, g^\alpha X^i_\alpha, \dot{v}^i) : \dot{v}^i = \dot{f}^\alpha X^i_\alpha + g^\beta f^\alpha X^j_\alpha \frac{\partial X^i_\beta}{\partial x^j},\ f^\alpha, g^\alpha, \dot{f}^\alpha \in \mathbb{R} \right\},$$

it follows that $s_{\mathcal{M}} \widehat{\mathcal{D}} = \widehat{\mathcal{D}}$ if and only if $X^j_\beta\, \partial X^i_\alpha / \partial x^j$ can be exchanged with $X^j_\alpha\, \partial X^i_\beta / \partial x^j$ plus an element of $\mathcal{D}$, i.e. if and only if $\mathcal{D}$ is involutive.

(3) $\Leftrightarrow$ (4): If $\widehat{\mathcal{D}} \subseteq s_{\mathcal{M}} T\mathcal{D}$ then $\widehat{\mathcal{D}} \subseteq s_{\mathcal{M}} (T\tau_{\mathcal{M}})^{-1} \mathcal{D} \cap s_{\mathcal{M}} T\mathcal{D} = s_{\mathcal{M}} \widehat{\mathcal{D}}$ because $\tau_{T\mathcal{M}} = T\tau_{\mathcal{M}} s_{\mathcal{M}}$ and $\widehat{\mathcal{D}} \subseteq (\tau_{T\mathcal{M}})^{-1} \mathcal{D}$. Since $s_{\mathcal{M}}$ is an involution, $\widehat{\mathcal{D}} \subseteq s_{\mathcal{M}} \widehat{\mathcal{D}}$ implies also $s_{\mathcal{M}} \widehat{\mathcal{D}} \subseteq \widehat{\mathcal{D}}$, i.e. $\widehat{\mathcal{D}} = s_{\mathcal{M}} \widehat{\mathcal{D}}$. Conversely, if $s_{\mathcal{M}} \widehat{\mathcal{D}} = \widehat{\mathcal{D}}$ then $\widehat{\mathcal{D}} = s_{\mathcal{M}} \widehat{\mathcal{D}} \subseteq s_{\mathcal{M}} T\mathcal{D}$ because $\widehat{\mathcal{D}} \subseteq T\mathcal{D}$.

Generally for injective immersions f and f', $f \subseteq f'$ means that $f = f'\psi$ for a smooth ψ. It is easy to show that $f \subseteq f'$ if and only if $Tf \subseteq Tf'$. Recall that if f, f' are (connected immersed) integral submanifolds of a distribution then $f \subseteq f'$ if and only if image $f \subseteq$ image f'.

If $\iota : \mathcal{N} \to \mathcal{M}$ is a (connected immersed) integral submanifold of $\mathcal{D}$, set $\widehat{\mathcal{N}} \equiv T\mathcal{N}$ and $\hat{\iota} \equiv T\iota$. Then $\hat{\iota} : \widehat{N} \to \mathcal{D}$ is an integral submanifold of $\widehat{\mathcal{D}}$ because image $T\hat{\iota} = T(\text{image}\, T\iota) \subseteq T\mathcal{D}$, and $T\tau_Q T\hat{\iota} = T(\tau_Q T\iota) = T\iota$ so image$(T\tau_Q T\hat{\iota}) \subseteq T\mathcal{D}$, hence image $T\hat{\iota} \subseteq T\mathcal{D} \cap T\tau_Q^{-1} \mathcal{D} = \widehat{\mathcal{D}}$. Thus the tangent of any integral submanifold of $\mathcal{D}$ is an integral submanifold of $\widehat{\mathcal{D}}$. Conversely, suppose $\hat{\iota} : \widehat{\mathcal{N}} \to \mathcal{D}$ is a (connected immersed) integral submanifold of $\widehat{\mathcal{D}}$. Then image$(T\tau_Q \hat{\iota}) \subseteq \mathcal{D}$, so $T\tau_Q \hat{\iota}$ factors through some integral submanifold $\iota : \mathcal{N} \to \mathcal{M}$ of $\mathcal{D}$ i.e. $T\tau_Q \hat{\iota} = \iota\psi$ for some ψ

([39], Theorem 1.62). By the above, $\boldsymbol{T}\iota$ is an integral submanifold of $\widehat{\mathcal{D}}$. For any $\hat{n} \in \widehat{\mathcal{N}}$, $\boldsymbol{T}\iota$ is defined at any element of $\mathcal{D}$ with base point $\iota(\psi(\hat{n})) = \boldsymbol{T}\tau_Q(\hat{\iota}(\hat{n}))$. Since $\hat{\iota}(\hat{n})$ is such an element of $\mathcal{D}$, the integral manifold $\boldsymbol{T}\iota$ passes through every point of $\mathcal{D}$ for which the integral manifold $\hat{\iota}$ is defined. Thus every integral manifold of $\widehat{\mathcal{D}}$ is contained in the tangent of an integral submanifold of $\mathcal{D}$.

If $\hat{\iota}$ is a maximal integral manifold of $\widehat{\mathcal{D}}$, then $\hat{\iota}$ is contained in the corresponding $\boldsymbol{T}\iota$ defined above, so by maximality $\hat{\iota} = \boldsymbol{T}\iota$. ι itself is maximal because if $\iota \subseteq \iota'$ then $\hat{\iota} = \boldsymbol{T}\iota \subseteq \boldsymbol{T}\iota'$ so $\boldsymbol{T}\iota = \boldsymbol{T}\iota'$, and then $\mathrm{image}(\iota) = \mathrm{image}(\iota')$ so $\iota = \iota'$. Thus every maximal integral submanifold of $\widehat{\mathcal{D}}$ is the tangent of a maximal integral submanifold of $\mathcal{D}$. If ι is a maximal integral submanifold of $\mathcal{D}$ and $\boldsymbol{T}\iota \subseteq \hat{\iota}$ then as above construct an integral submanifold ι' of $\mathcal{D}$ such that $\hat{\iota} \subseteq \boldsymbol{T}\iota'$. Then $\boldsymbol{T}\iota \subseteq \boldsymbol{T}\iota'$ so $\iota \subseteq \iota'$ and hence $\iota = \iota'$ by maximality of ι. Thus $\hat{\iota} \subseteq \boldsymbol{T}\iota' = \boldsymbol{T}\iota$ so $\boldsymbol{T}\iota = \hat{\iota}$, which shows that $\boldsymbol{T}\iota$ is maximal. Thus the tangent of every maximal integral submanifold of $\mathcal{D}$ is a maximal integral submanifold of $\widehat{\mathcal{D}}$. □

The distribution $\widehat{\mathcal{D}}$, or objects similar to it or diffeomorphic to it, occur ubiquitously. For example, the image of $\widehat{\mathcal{D}}$ under the derivative of the Legendre transformation is the distribution H of [4]. In [14], that H is pulled back to the tangent bundle to obtaining $\widehat{\mathcal{D}}$ but it is again denoted H. It is defined as above in [36], and it is the restriction of the distribution obtained as the complete lifts of vector fields in $\mathcal{D}$ discussed by [20].

If X is a vector field then write $X \in \mathcal{D}$ when $X(p) \in \mathcal{D}$ for all p. The following lemma is covered by [37] but is included here for convenience.

LEMMA 2.2. *If X is a vector field such that $[X, Y] \in \mathcal{D}$ for all $Y \in \mathcal{D}$, and if the flow of X is F_t, then $\boldsymbol{T}F_t(\mathcal{D}) \subseteq \mathcal{D}$.*

Proof: It suffices to show that the flow $\boldsymbol{T}F_t$ is tangent to $\boldsymbol{T}\mathcal{D}$. If $v_p \in \mathcal{D}$ then picking a vector field $V \in \mathcal{D}$ with $V(p) = v_p$ gives

$$\begin{aligned}
\left.\frac{d}{dt}\right|_{t=0} \boldsymbol{T}F_t\, v_p &= \left.\frac{d}{dt}\right|_{t=0} \boldsymbol{T}F_t\, V\big(F_{-t}F_t(p)\big) \\
&= \left.\frac{d}{dt}\right|_{t=0} \boldsymbol{T}F_t\, V\big(F_{-t}(p)\big) + \left.\frac{d}{dt}\right|_{t=0} V\big(F_t(p)\big) \\
&= -\,\mathrm{vert}_{v_p}[X, V](p) + \left.\frac{d}{dt}\right|_{t=0} V\big(F_t(p)\big)
\end{aligned}$$

which is in $\boldsymbol{T}\mathcal{D}$. □

The condition $[X, Y] \in \mathcal{D}$ for all $Y \in \mathcal{D}$ is called $\mathcal{D}$ *invariance under* X by [24]. They show that if $\mathcal{D}$ is involutive and is associated to a foliation then X is projectable to the quotient if and only if it is $\mathcal{D}$-invariant. Lemma 2.2 flips this, showing that $\mathcal{D}$ is X-invariant (in a flow sense) if X is $\mathcal{D}$ invariant, regardless of whether $\mathcal{D}$ is involutive.

An *r-form on* $\mathcal{D}$ is a smooth section of $\Omega^r\mathcal{D}$. Let α be an r-form in $\mathcal{D}$ and $[X, Y] \in \mathcal{D}$ for all $Y \in \mathcal{D}$. By Lemma 2.2, the r-form $\boldsymbol{L}_X\alpha$ in $\mathcal{D}$ can be defined by the usual Lie formula

$$\boldsymbol{L}_X\alpha \equiv \left.\frac{d}{dt}\right|_{t=0} F_t^*\alpha,$$

and one has the usual expression

$$\boldsymbol{L}_X\alpha(Y_1, \cdots, Y_r) = X\big(\alpha(Y_1, \cdots, Y_r)\big) - \sum_{i=1}^{r} \alpha(Y_1, \cdots, [X, Y_i], \cdots, Y_r).$$

For vector fields $Y_0, \ldots, Y_r \in \mathcal{D}$ which have pairwise bracket in $\mathcal{D}$, define

$$\begin{aligned} \boldsymbol{d}\alpha(Y_0, \ldots, Y_r) \equiv &\sum_{i=0}^{r} Y_i\big(\alpha(Y_0, \ldots, \hat{Y}_i, \ldots, Y_r)\big) \\ &+ \sum_{i<j}^{r} \alpha([Y_i, Y_j], Y_0, \ldots, \hat{Y}_i, \ldots, \hat{Y}_j, \ldots, Y_r). \end{aligned}$$

If α is the restriction to $\mathcal{D}$ of a form $\hat{\alpha}$ on $\mathcal{M}$ then $\boldsymbol{d}\alpha$ as above may be computed by insertion of the vector fields $Y_0, \ldots, Y_r$ into $\boldsymbol{d}\hat{\alpha}$ as that is usually defined. If $\mathcal{D}_0$ is a subbundle of $\mathcal{D}$ such that $[X, Y] \in \mathcal{D}$ for all $X, Y \in \mathcal{D}_0$ then the definition of $\boldsymbol{d}\alpha$ well defines it as a form on $\mathcal{D}_0$. In any case, if $Y_0, \ldots, Y_r$ have pairwise bracket in $\mathcal{D}$ then

$$\boldsymbol{L}_X\alpha(Y_1, \ldots, Y_r) = \boldsymbol{d}(\boldsymbol{i}_X\alpha)(Y_1, \ldots, Y_r) + \boldsymbol{i}_X(\boldsymbol{d}\alpha)(Y_1, \ldots, Y_r).$$

Suppose that $\mathcal{E}$ is a vector bundle and $\nu : \boldsymbol{T}\mathcal{M} \to \mathcal{E}$ is a vector bundle map such that $\mathcal{D} \subseteq \ker\nu$. For vector fields $X, Y \in \mathcal{D}$, $\nu[X, Y]$ in local coordinates is

$$(\nu[X, Y])^a = \nu_i^a\left(X^j\frac{\partial Y^i}{\partial x^j} - Y^j\frac{\partial X^i}{\partial x^j}\right). \tag{4}$$

$\nu_i^a X^i = 0$ and $\nu_i^a Y^i = 0$ since X and Y are in $\mathcal{D}$ and ν annihilates $\mathcal{D}$. Differentiating gives

$$\nu_i^a\frac{\partial X^i}{\partial x^j} = -X^i\frac{\partial \nu_i^a}{\partial x^j}, \qquad \nu_i^a\frac{\partial Y^i}{\partial x^j} = -Y^i\frac{\partial \nu_i^a}{\partial x^j}$$

and substituting into (4) gives

$$(\nu[X, Y])^a = -\frac{\partial \nu_i^a}{\partial x^j}Y^iX^j + \frac{\partial \nu_i^a}{\partial x^j}X^iY^j = \left(\frac{\partial \nu_i^a}{\partial x^j} - \frac{\partial \nu_j^a}{\partial x^i}\right)X^iY^j.$$

In particular, there is a unique $\mathcal{E}$-valued *curvature two-form* $\Delta_{\mathcal{D}}$ on $\boldsymbol{T}\mathcal{M}$ such that, for all vector fields $X, Y \in \mathcal{D}$,

$$\Delta_{\mathcal{D}}(X, Y) = -\nu[X, Y],$$

and in coordinates

$$\Delta_{\mathcal{D}}^a = -\left(\frac{\partial \nu_i^a}{\partial x^j} - \frac{\partial \nu_j^a}{\partial x^i}\right). \tag{5}$$

If $\ker \nu = \mathcal{D}$ then $\mathcal{D}$ is involutive if and only if $\Delta_{\mathcal{D}} = 0$. One can always choose $\mathcal{E} = T\mathcal{M}/\mathcal{D}$ and thus there is always an $\mathcal{E} = T\mathcal{M}/\mathcal{D}$-valued curvature two-form that is zero if and only if $\mathcal{D}$ is involutive. It is also noted in [6] that the notion of curvature is available as an $T\mathcal{M}/\mathcal{D}$-valued form without a bundle structure or connection.

A special case occurs where the distribution is explicitly defined as the image of a projection $\kappa : T\mathcal{M} \to \mathcal{M}$. Then $\mathcal{D} = \ker \nu$ where $\nu = \mathbf{1}_{T\mathcal{M}} - \kappa$ and

$$\kappa[\kappa X, \kappa Y] - [\kappa X, \kappa Y] = \Delta_\kappa(X, Y),$$

where $\Delta_\kappa(X, Y) \equiv \Delta_{\mathcal{D}}(\kappa X, \kappa Y)$. For a function f, after defining $\boldsymbol{d}_\kappa f \equiv \kappa^* \boldsymbol{d} f$, the computation

$$\begin{aligned}
\boldsymbol{d}_\kappa^2 f(X, Y) &= \kappa X\big(\boldsymbol{d}_\kappa f(\kappa Y)\big) - \kappa Y\big(\boldsymbol{d}_\kappa f(\kappa X)\big) - \boldsymbol{d}_\kappa f([\kappa X, \kappa Y]) \\
&= \kappa X\big(\boldsymbol{d} f(\kappa Y)\big) - \kappa Y\big(\boldsymbol{d} f(\kappa X)\big) - \boldsymbol{d} f(\kappa[\kappa X, \kappa Y]) \\
&= \kappa X\big(\boldsymbol{d} f(\kappa Y)\big) - \kappa Y\big(\boldsymbol{d} f(\kappa X)\big) - \boldsymbol{d} f\big([\kappa X, \kappa Y] + \Delta_\kappa(X, Y)\big) \\
&= \boldsymbol{dd} f(\kappa X, \kappa Y) - \langle \boldsymbol{d} f, \Delta_\kappa(X, Y)\rangle \\
&= -\langle \boldsymbol{d} f, \Delta_\kappa(X, Y)\rangle
\end{aligned}$$

gives $\boldsymbol{d}_\kappa^2 f = -\langle \boldsymbol{d} f, \Delta_\kappa\rangle$, and $\boldsymbol{d}_\kappa^2 = 0$ on functions if and only if $\mathcal{D}$ is involutive. The vectors in the distribution $\mathcal{D}$ are called *horizontal* whereas those in $\ker \kappa = \operatorname{Image} \nu$ are called *vertical*. Both ν and Δ_κ take values in $\ker \kappa$; these are *vertical-valued forms*.

Although not needed here, there is a calculus of vertical-valued forms which includes the formula $\boldsymbol{d}_\kappa^2 = -\langle \boldsymbol{d} f, \Delta_\kappa\rangle$. This is established in the context of vector bundle Ehresmann connections in [25] but it is available for any distribution which is the image of a projection, i.e. for any distribution with an explicit complement. The *exterior covariant derivative* $\boldsymbol{d}_\kappa \alpha$ of a vertical-valued r-form α can be defined by

$$\begin{aligned}
\boldsymbol{d}_\kappa \alpha(X_0, \cdots, X_r) \equiv{} & \sum_{i=0}^{k} (-1)^i \nu[\kappa X_i, \alpha(\kappa X_0, \cdots, \widehat{\kappa X_i}, \cdots, \kappa X_r)] \\
& + \sum_{i<j} (-1)^{i+j} \alpha([\kappa X_i, \kappa X_j], \kappa X_0, \cdots, \widehat{\kappa X_i}, \cdots, \widehat{\kappa X_j}, \cdots, \kappa X_r).
\end{aligned}$$

Then

$$\boldsymbol{d}_\kappa \nu(X, Y) = \nu[\kappa X, \nu(\kappa Y)] - \nu[\kappa Y, \nu(\kappa X)] - \nu[\kappa X, \kappa Y] = \Delta_\kappa(X, Y)$$

and there is a Bianchi identity because

$$\begin{aligned}
\boldsymbol{d}_\kappa \Delta_\kappa(X, Y, Z) &= \nu[\kappa X, \Delta_\kappa(\kappa Y, \kappa Z)] - \Delta_\kappa([\kappa X, \kappa Y], \kappa Z) + \circlearrowleft \\
&= -\nu\big[\kappa X, \nu[\kappa Y, \kappa Z]\big] + \nu\big[\kappa[\kappa X, \kappa Y], \kappa Z\big] + \circlearrowleft \\
&= -\nu\big[\kappa X, \nu[\kappa Y, \kappa Z]\big] + \nu\big[\kappa[\kappa Y, \kappa Z], \kappa X\big] + \circlearrowleft \\
&= -\nu\big[\kappa X, [\kappa Y, \kappa Z]\big] + \circlearrowleft \\
&= 0.
\end{aligned}$$

After the definition of the curvature in terms of the Lie bracket, the Bianchi identity is simply an expected reflection of the Jacobi identity.

A fact that will be used later: if $c(s,t) \in \mathcal{M}$ is C^2, $\partial c/\partial s(0,0) \in \mathcal{D}$, and $\partial c/\partial t(0,0) \in \mathcal{D}$, then

$$s_{\mathcal{M}}\, \boldsymbol{T}\kappa\, \frac{\partial^2 c}{\partial s \partial t}(0,0) - \boldsymbol{T}\kappa\, \frac{\partial^2 c}{\partial t \partial s}(0,0) = -\operatorname{vert}_{\partial c/\partial s} \Delta_\kappa \left(\frac{\partial c}{\partial s}, \frac{\partial c}{\partial t} \right) \Bigg|_{s=t=0}. \tag{6}$$

Indeed, in coordinates

$$\begin{aligned}
\boldsymbol{T}\kappa\, \frac{\partial^2 x^i}{\partial s \partial t}(0,0) &= \frac{\partial}{\partial s} \left(x^i(s,t), \kappa^i_j\big(x(s,t)\big) \frac{\partial x^j}{\partial t} \right) \Bigg|_{s=t=0} \\
&= \left(x^i, \kappa^i_j \frac{\partial x^j}{\partial t}, \frac{\partial x^i}{\partial s}, \frac{\partial \kappa^i_j}{\partial x^k} \frac{\partial x^k}{\partial s} \frac{\partial x^j}{\partial t} + \kappa^i_j \frac{\partial^2 x^j}{\partial s \partial t} \right) \Bigg|_{s=t=0} \\
&= \left(x^i, \frac{\partial x^i}{\partial t}, \frac{\partial x^i}{\partial s}, \frac{\partial \kappa^i_j}{\partial x^k} \frac{\partial x^k}{\partial s} \frac{\partial x^j}{\partial t} + \kappa^i_j \frac{\partial^2 x^j}{\partial s \partial t} \right) \Bigg|_{s=t=0},
\end{aligned}$$

so that

$$\begin{aligned}
& s_{\mathcal{M}}\, \boldsymbol{T}\kappa\, \frac{\partial^2 x^i}{\partial s \partial t}(0,0) - \boldsymbol{T}\kappa\, \frac{\partial^2 x^i}{\partial t \partial s}(0,0) \\
&\quad = \left(x^i, \frac{\partial x^i}{\partial s}, 0, \frac{\partial \kappa^i_j}{\partial x^k} \frac{\partial x^k}{\partial s} \frac{\partial x^j}{\partial t} - \frac{\partial \kappa^i_j}{\partial x^k} \frac{\partial x^j}{\partial s} \frac{\partial x^k}{\partial t} \right) \Bigg|_{s=t=0} \\
&\quad = \left(x^i, \frac{\partial x^i}{\partial s}, 0, \left(\frac{\partial v^i_j}{\partial x^k} - \frac{\partial v^i_k}{\partial x^j} \right) \frac{\partial x^j}{\partial s} \frac{\partial x^k}{\partial t} \right) \Bigg|_{s=t=0},
\end{aligned}$$

which in view of (5) is as required.

3. Constrained Lagrangian systems

Suppose f is a function (the *objective*) on a manifold $\mathcal{M}$ and $\mathcal{N}$ is a submanifold (the *constraint*) of $\mathcal{M}$. The *constrained variational problem* defined by f and $\mathcal{N}$ is the finding of the critical points of $f|\mathcal{N}$, i.e. the solution of the equations $\boldsymbol{d}f(\boldsymbol{T}\mathcal{N}) = 0$. A more general problem is to find solutions to the equation $\boldsymbol{d}f(\mathcal{A}|\mathcal{N}) = 0$ where $\mathcal{A}$ is some *a priori* given distribution which is *different* from $\boldsymbol{T}\mathcal{N}$. Both problems are variational in the sense that each zeros are directional derivatives of an objective.

DEFINITION 3.1. A *skew variational problem* is a tuple $(\mathcal{M}, \mathcal{N}, \mathcal{D}, f)$ where $\mathcal{M}$ and $\mathcal{N}$ are manifolds, $\mathcal{D} \subseteq \boldsymbol{T}\mathcal{M}|\mathcal{N}$, and $f : \mathcal{M} \to \mathbb{R}$ is differentiable. The *solutions* or *critical points* are the $m \in \mathcal{N}$ such that

$$\boldsymbol{d}f(m)v = 0 \qquad \text{for all} \quad v \in \mathcal{D} \cap \boldsymbol{T}_m\mathcal{M}.$$

A skew variational problem is *variational* if $\mathcal{D} \subseteq \boldsymbol{T}\mathcal{N}$ and $\mathcal{D}$ is an involutive distribution of $\mathcal{N}$. It is an *ordinary critical point problem* if $\mathcal{D} = \boldsymbol{T}\mathcal{N}$.

Given a skew variational problem $(\mathcal{M}, \mathcal{N}, \mathcal{D}, f)$ where $\mathcal{D}$ has fiber dimension d, there will typically be $(\dim N - d)$-dimensional submanifolds of critical points, and this is true if the problem is variational as well. If the skew problem is variational, then it is also a family of ordinary critical point problems where the constraints are the leaves of $\mathcal{D}$, so variational problems are naturally ordinary critical point problems which are parameterized by the quotient $\mathcal{N}/\mathcal{D}$. This is relevant for constrained Lagrangian systems because the fixed boundary constraint for nonholonomic systems is implemented in the variations on the curves, i.e. in the analogue of $\mathcal{D}$, while for holonomic systems it is usually implemented as a constraint on curves, i.e. in part of an analogue of $\mathcal{N}$. When regarding holonomic systems as special cases of nonholonomic ones, the explicit fixed-endpoint condition emerges from leaves of $\mathcal{D}$. It is a matter of choice whether the leaves are imposed as a constraint, leading to a parameterized set of an ordinary critical point problems, or not, leading to a variational problem.

As was just anticipated, formally, a nonholonomic system, defined by a Lagrangian $L : \boldsymbol{T}\mathcal{Q} \to \mathbb{R}$ and a constraint distribution $\mathcal{D}$, is variational if and only if $\mathcal{D}$ is involutive: Let $\mathcal{N}_{[a,b]}$ be the set of curves in $\mathcal{Q}$, defined on $[a, b]$, that satisfy the constraint $\mathcal{D}$, defined by

$$\mathcal{N}_{[a,b]} \equiv \left\{ q(t) : q'(t) \in \mathcal{D} \right\},$$

and let $\mathcal{D}_{[a,b]}$ be the set of fixed-endpoint variations along curves in $\mathcal{N}_{[a,b]}$, defined by

$$\mathcal{D}_{[a,b]} \equiv \left\{ \delta q(t) : \tau_{\mathcal{Q}}\, \delta q \in \mathcal{N}_{[a,b]},\ \delta q(t) \in \mathcal{D},\ \delta q(a) = 0,\ \delta q(b) = 0 \right\}.$$

The Lagrange–d'Alembert principle seeks curves $q(t)$ for which the action on $\mathcal{N}_{[a,b]}$ differentiates to zero in the directions defined by $\mathcal{D}_{[a,b]}$. This principle corresponds to a constrained variational principle when $\mathcal{D}_{[a,b]}$ is a subset of the tangent space of $\mathcal{N}_{[a,b]}$. *If it is a subset*, then recalling that $s_{\mathcal{Q}} : \boldsymbol{TT}\mathcal{Q} \to \boldsymbol{TT}\mathcal{Q}$ denotes the canonical involution,

$$\begin{aligned}
\boldsymbol{T}\mathcal{N}_{[a,b]} &\subseteq \left\{ \left. \frac{\partial}{\partial \epsilon} \right|_{\epsilon=0} q_\epsilon(t) : q'_\epsilon(t) \in \mathcal{D} \right\} \\
&\subseteq \left\{ \left. \frac{\partial}{\partial \epsilon} \right|_{\epsilon=0} q_\epsilon(t) : q'_0(t) \in \mathcal{D},\ s_{\mathcal{Q}} \left. \frac{\partial}{\partial t} \frac{\partial}{\partial \epsilon} \right|_{\epsilon=0} q_\epsilon(t) \in \boldsymbol{T}\mathcal{D} \right\} \\
&= \left\{ \delta q(t) : \tau_{\mathcal{Q}}\, \delta q \in \mathcal{N}_{[a,b]},\ \delta q'(t) \in s_{\mathcal{Q}} \boldsymbol{T}\mathcal{D} \right\},
\end{aligned}$$

whereas, recalling that $\widehat{\mathcal{D}} = (\boldsymbol{T}\tau_{\mathcal{Q}})^{-1}\mathcal{D} \cap \boldsymbol{T}\mathcal{D}$,

$$\begin{aligned}
\mathcal{D}_{[a,b]} &= \left\{ \delta q(t) : \boldsymbol{T}\tau_{\mathcal{Q}}\, \delta q'(t) \in \mathcal{D},\ \delta q'(t) \in \boldsymbol{T}\mathcal{D},\ \delta q(a) = 0,\ \delta q(b) = 0 \right\} \\
&= \left\{ \delta q(t) : \tau_{\mathcal{Q}}\, \delta q \in \mathcal{N}_{[a,b]},\ \delta q'(t) \in \widehat{\mathcal{D}},\ \delta q(a) = 0,\ \delta q(b) = 0 \right\}.
\end{aligned}$$

By comparison, $\mathcal{D}_{[a,b]} \subseteq \boldsymbol{T}\mathcal{N}_{[a,b]}$ implies $\widehat{\mathcal{D}} \subseteq s_{\mathcal{Q}} \boldsymbol{T}\mathcal{D}$, which by Lemma 2.1 implies that $\mathcal{D}$ is involutive. *Conversely*, if $\mathcal{D}$ is involutive then $\mathcal{N}_{[a,b]}$ is the set of curves on $[a, b]$ in leaves of $\mathcal{D}$ and this has tangent space of the curves in the tangent space of those leaves, from which $\mathcal{D}_{[a,b]} \subseteq T\mathcal{N}_{[a,b]}$ follows. Futhermore, $\mathcal{D}_{[a,b]}$ is formally integrable, because its leaves are the curves in the leaves of $\mathcal{D}$ with specific fixed endpoints, so the Lagrangian system is, formally, variational.

The critical curves of S under the constraint $q'(t) \in \mathcal{D}$ are by definition the vakanomic evolution curves. The argument just presented implies that vakanomic mechanics and nonholonomic mechanics are equivalent if the constraints are linear and integrable, thus recovering Proposition 2.8 of [21]. Their context is affine constraints, and they obtain equivalence assuming the constraints to be "affine integrable", by a direct comparison of the equations of motion. By their definition, affine integrable constraints are actually linear, so their result is also one of linear constraints.

Two distinct aspects occur in a nonholonomic system. In one aspect, $\mathcal{D}$ determines a constraint, which is subset of the space of curves on $\mathcal{Q}$: one seeks *within* $\mathcal{N}_{[a,b]}$ a kind of critical point. In another aspect, $\mathcal{D}$ determines the *directions* $\mathcal{D}_{[a,b]}$ in which the derivative vanishes. The object that allows $\mathcal{D}$ to serve the two roles simultaneously is the derivative; it is this that in the definition of $\mathcal{N}_{[a,b]}$ allows the construction of $\mathcal{N}_{[a,b]}$ from $\mathcal{D}$.

It is useful to recite a result of [26], which formalizes the identification of the Lagrange one-form with the variation of the action.

THEOREM 3.1. *Given a C^k Lagrangian L, $k \geq 2$, there exists a unique C^{k-2} mapping $\delta L : \ddot{\mathcal{Q}} \to T^*\mathcal{Q}$, defined on the second-order submanifold*

$$\ddot{\mathcal{Q}} \equiv \{q''(0) : q(t) \text{ a } C^2 \text{ curve in } \mathcal{Q}\}$$

of $TT\mathcal{Q}$, and a unique C^{k-1} 1-form θ_L on $T\mathcal{Q}$, such that, for all C^2 variations $q_\epsilon(t)$,

$$\boldsymbol{d}S\big(q(t)\big) \cdot \delta q(t) = \int_a^b \delta L\big(q''(t)\big) \cdot \delta q(t)\, dt + \theta_L\big(q'(t)\big) \cdot \widehat{\delta q}\Big|_a^b, \tag{7}$$

where

$$\delta q(t) \equiv \frac{\partial}{\partial \epsilon}\Big|_{\epsilon=0} q_\epsilon(t), \qquad \widehat{\delta q}(t) \equiv \frac{\partial}{\partial \epsilon}\Big|_{\epsilon=0} \frac{\partial}{\partial t}\Big|_{t=0} q_\epsilon(t).$$

θ_L so defined is called the Lagrange 1-form.

The point of Theorem 3.1 is that the variational principle identifies δL and θ_L directly. In coordinates, $\mathcal{Q} = \{(q^i, \dot{q}^i, \dot{q}^i, \ddot{q}^i)\}$ and

$$\delta L = \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right)\boldsymbol{d}q^i = \left(\frac{\partial L}{\partial q^i} - \frac{\partial^2 L}{\partial \dot{q}^i \partial q^j}\dot{q}^j - \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}\ddot{q}^j\right)\boldsymbol{d}q^i.$$

A direct demonstration of the covariance of δL can be found in [33].

In this work the variational principle drives the development towards the relevant differential-geometric objects, so these cannot be used to define the vector field for the time evolution, which rather is required at the outset. Recall [1] that the *second fiber derivative of L* is

$$\boldsymbol{F}^2L(v_q)(w_q, \tilde{w}_q) = \boldsymbol{D}^2(L|\boldsymbol{T}_q\mathcal{Q})(w_q, \tilde{w}_q).$$

DEFINITION 3.2. A Lagrangian L is *$\mathcal{D}$-regular* if, for all $v_q \in \mathcal{D}$, the bilinear form $\boldsymbol{F}^2L(v_q)|\mathcal{D}_q \times \mathcal{D}_q$ is nonsingular.

THEOREM 3.2. *If L is a C^k $\mathcal{D}$-regular Lagrangian, $k \geq 2$, and $\mathcal{D}$ is C^{k-2}, then there is a unique C^{k-2} second-order vector field $Y_{\delta L}$ on $\mathcal{D}$ such that $\delta L \circ Y_{\delta L} \in \operatorname{ann}\mathcal{D}$. Moreover, $Y_{\delta L}$ is in $\mathcal{K}_{\mathcal{D}} \equiv \widehat{\mathcal{D}}$.*

Proof: It is sufficient to show *local* existence and uniqueness of $Y_{\delta L}$. Suppose that the vector fields X_a are fiberwise a basis of $\mathcal{D}$, giving coordinates (q^i, f^a) on $\mathcal{D}$. Any vector field on $\mathcal{D}$ that is second order is uniquely of the form

$$Y = f^a X_a^i \frac{\partial}{\partial q^i} + \left(f^a f^b X_b^j \frac{\partial X_a^i}{\partial q^j} + \dot{f}^a X_a^i \right) \frac{\partial}{\partial \dot{q}^i}.$$

Substitution of this into $\delta L \circ X = 0$ and insertion of X_b gives

$$\left[\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} X_a^i X_b^j \right] \dot{f}^a = \left(\frac{\partial L}{\partial q^i} - \frac{\partial^2 L}{\partial \dot{q}^i \partial q^j} f^a X_a^j - \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} f^a f^c X_c^k \frac{\partial X_a^j}{\partial q^k} \right) X_b^i,$$

$\mathcal{D}$-regularity is exactly what is required to establish this as a nonsingular linear equation for $\dot{f}^a$ and thus to determine $\dot{f}^a$ as functions of (q^i, f^a). □

The change in notation to $\mathcal{K}_{\mathcal{D}}$ from $\widehat{\mathcal{D}}$ anticipates the context of semi-Hamiltonian systems, where there will be an analogue of $\widehat{\mathcal{D}}$ but none of $\mathcal{D}$. The kinematic meaning of the distribution $\mathcal{K}_{\mathcal{D}}$ is simple: it is the bundle of possible infinitesimal motions in the phase space $\mathcal{D}$ of states of the nonholonomic system, absent the first-order condition. Indeed, $\mathcal{K}_{\mathcal{D}}$ is by definition the set of derivatives of all curves $v(t)$ that are in $\mathcal{D}$ and whose base integral curves have derivative in $\mathcal{D}$. The fiber dimension of $\mathcal{K}_{\mathcal{D}}$ is even because these two conditions account for equal halves.

The use of "Y" for the vector fields of nonholonomic mechanics appears in [35, 36]. This neatly avoids conflict with the already used notation X_f for the Hamiltonian vector field of f in contexts where nonholonomic systems coexist with a symplectic context.

The restriction of a positive-definite symmetric bilinear form to a subspace is also positive definite, and hence nonsingular. So, as is observed in many works on nonholonomic mechanics, in the usual situation where the Lagrangian L is of the form kinetic plus potential, $\boldsymbol{F}^2L$ is equal to the kinetic energy metric and L is $\mathcal{D}$-regular for any $\mathcal{D}$.

Suppose that $\mathcal{D}$ is the image of a projection κ, and define the *restricted Lagrangian* L_c by $L_c = L \circ \kappa$. In [6], Theorem 2.4, there is a characterization of the Lagrange–d'Alembert equations in terms of the curvature of $\mathcal{D}$ and δL_c. This follows simply, without coordinates, directly from the variational principle and equation (6), as follows. If $q(t)$ is such that $q'(t) \in \mathcal{D}$ and q_ϵ is a variation of $q(t)$, then

$$\delta q(t) \equiv \left. \frac{\partial}{\partial \epsilon} \right|_{\epsilon=0} q_\epsilon(t) \in \mathcal{D},$$

then

$$
\begin{aligned}
&\frac{d}{d\epsilon}\bigg|_{\epsilon=0}\int_a^b L_c\left(\frac{\partial q_\epsilon}{\partial t}\right)dt \\
&\quad=\int_a^b \boldsymbol{d}L\left(\boldsymbol{T}\kappa\,\frac{\partial^2 q_\epsilon}{\partial\epsilon\partial t}\right)\bigg|_{\epsilon=0} \\
&\quad=\int_a^b \boldsymbol{d}L\left(s_Q\,\boldsymbol{T}\kappa\,\frac{\partial^2 q_\epsilon}{\partial t\partial\epsilon}+\operatorname{vert}_{\partial q_\epsilon/\partial t}\,\Delta_\kappa\left(\frac{\partial q_\epsilon}{\partial t},\frac{\partial q_\epsilon}{\partial\epsilon}\right)\right)\bigg|_{\epsilon=0} \\
&\quad=\int_a^b \boldsymbol{d}L\left(s_Q\,\frac{\partial}{\partial t}\,\kappa\,\frac{\partial q_\epsilon}{\partial\epsilon}\bigg|_{\epsilon=0}\right)+\boldsymbol{d}L\left(\operatorname{vert}_{\partial q_\epsilon/\partial t}\,\Delta_\kappa\left(\frac{\partial q_\epsilon}{\partial t},\frac{\partial q_\epsilon}{\partial\epsilon}\right)\right)\bigg|_{\epsilon=0} \\
&\quad=\int_a^b \boldsymbol{d}L\left(s_Q\,\frac{\partial}{\partial t}\,\frac{\partial q_\epsilon}{\partial\epsilon}\right)\bigg|_{\epsilon=0}+\boldsymbol{F}L\left(\frac{\partial q_\epsilon}{\partial t}\right)\Delta_\kappa\left(\frac{\partial q_\epsilon}{\partial t},\frac{\partial q_\epsilon}{\partial\epsilon}\right)\bigg|_{\epsilon=0} \\
&\quad=\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}\int_a^b L\left(\frac{\partial q_\epsilon}{\partial t}\right)+\boldsymbol{F}L\big(q'(t)\big)\Delta_\kappa\big(q'(t),\delta q(t)\big)
\end{aligned}
$$

from which

$$\delta L_c\big(q''(t)\big)\delta q(t)=\delta L\big(q''(t)\big)\delta q(t)+\boldsymbol{F}L\big(q'(t)\big)\Delta_\kappa\big(q'(t),\delta q(t)\big). \tag{8}$$

The Lagrange–d'Alembert equations hold if and only if $q'(t)\in\mathcal{D}$ and δL vanishes for all $\delta q\in\mathcal{D}$. From (8), this is equivalent to $q'(t)\in\mathcal{D}$ and the Lagrange–d'Alembert equations of the restricted Lagrangian equal the curvature-related term on the right. The restricted Lagrangian is not regular except if $\mathcal{D}=\boldsymbol{T}Q$.

4. Variational development

In [26] the variational development of the geometry of unconstrained Lagrangian systems $L:\boldsymbol{T}Q\to\mathbb{R}$ is as follows. One fixes t and identifies the solution curves over the interval $[0,t]$ with the integral curve $t\mapsto F_t(v_q)$, where F_t is the Lagrangian flow. The pull back of the action functional through this identification is $S_t:\boldsymbol{T}Q\to\mathbb{R}$, where

$$S_t(v_q)=\int_0^t L\big(q'(s)\big)\,ds,\qquad q(s)=F_s(v_q).$$

Equation (7) pulls back to

$$\boldsymbol{d}S_t(v_q)w_{v_q}=\theta_L\big(F_t(v_q)\big)\cdot\frac{d}{d\epsilon}\bigg|_{\epsilon=0}F_t(v_q+\epsilon w_{v_q})-\theta_L(v_q)\cdot w_{v_q},$$

where $v_q+\epsilon w_{v_q}$ symbolically denotes any curve in ϵ at v_q with derivative w_{v_q} at $\epsilon=0$. Thus

$$\boldsymbol{d}S_t=F_t^*\theta_L-\theta_L. \tag{9}$$

The exterior derivative of this equation gives

$$0=\boldsymbol{dd}S_t=\boldsymbol{d}(F_t^*\theta_L)-\boldsymbol{d}\theta_L=-F_t^*\omega_L+\omega_L,$$

and so $F_t^*\omega_L = \omega_L$, i.e. the Lagrangian flow is symplectic. If L is invariant under lifts of an action of a Lie group $\mathcal{G}$ on $\mathcal{Q}$, then S_t is $\mathcal{G}$-invariant. Inserting the infinitesimal generator $\xi_{T\mathcal{Q}}$ into (9) gives

$$0 = \boldsymbol{i}_{\xi_{T\mathcal{Q}}}\boldsymbol{d}S_t = \boldsymbol{i}_{\xi_{T\mathcal{Q}}}(F_t^*\theta_L - \theta_L) = F_t^*(\boldsymbol{i}_{\xi_{T\mathcal{Q}}}\theta_L) - \boldsymbol{i}_{\xi_{T\mathcal{Q}}}\theta_L,$$

i.e. the flow preserves the momentum $J_\xi = \boldsymbol{i}_{\xi_{T\mathcal{Q}}}\theta_L$. The differential-geometric equations for the Lagrangian vector field $Y_{\delta L}$ follow because one t derivative of (9) at $t = 0$ gives $\boldsymbol{d}L = \boldsymbol{L}_{Y_{\delta L}}\theta_L$, so that

$$\boldsymbol{i}_{Y_{\delta L}}\omega_L = -\boldsymbol{i}_{Y_{\delta L}}\boldsymbol{d}\theta_L = -\boldsymbol{L}_{Y_{\delta L}}\theta_L + \boldsymbol{d}(\boldsymbol{i}_{Y_{\delta L}}\theta_L) = \boldsymbol{d}(\boldsymbol{i}_{Y_{\delta L}}\theta_L - L) = \boldsymbol{d}E,$$

where $E \equiv \boldsymbol{i}_{Y_{\delta L}}\theta_L - L$.

If the action is pulled back by the flow of an arbitrary second-order vector field Y, rather that the Lagrangian evolution vector field $Y_{\delta L}$, then in exactly the same way as just above there results

$$\boldsymbol{d}L(w_{v_q}) = \boldsymbol{L}_Y\theta_L(w_{v_q}) + \delta L\big(Y(v_q)\big)\big(\boldsymbol{T}\tau_{\mathcal{Q}}(w_{v_q})\big)$$

for any $w_{v_q} \in T\mathcal{Q}$. Again as above

$$\boldsymbol{L}_Y\theta_L = \boldsymbol{i}_Y\boldsymbol{d}\theta_L + \boldsymbol{d}(\boldsymbol{i}_Y\theta_L) = -\boldsymbol{i}_Y\omega_L + \boldsymbol{d}(E + L)$$

so

$$\omega_L(y_{v_q}, w_{v_q}) - \boldsymbol{d}E(w_{v_q}) = \delta L(y_{v_q})\boldsymbol{T}\tau_{\mathcal{Q}}w_{v_q}, \qquad y_{v_q} \in \ddot{\mathcal{Q}}, w_{v_q} \in T\mathcal{Q}, \tag{10}$$

an equation that explains the relationship between δL, ω_L, and E.

The Lagrange–d'Alembert equations (1) use the same action functional as the holonomic equations except that the derivative $\boldsymbol{d}S$ vanishes only for variations in $\mathcal{D}$. The nonholonomic phase space $\mathcal{D}$ provides the analogous identification of the solution curves, i.e. $v_q \in \mathcal{D}$ is identified with $F_t(v_q)$, where F_t is the nonholonomic flow. Under this identification

$$\boldsymbol{d}S\big(q(t)\big) \cdot w_{v_q} = (F_t^*\theta_L - \theta_L)w_{v_q} + \int_0^t \delta L\big(q''(s)\big) \cdot \delta q\, ds, \tag{11}$$

where δq is the *projected variation*

$$\delta q = \boldsymbol{T}\tau_{\mathcal{Q}}\,\boldsymbol{T}F_t(w_{v_q}) = \boldsymbol{T}(\tau_{\mathcal{Q}}\,F_t)(w_{v_q}). \tag{12}$$

Suppose $\mathcal{D}$ is involutive, corresponding to a holonomic system, and fix an initial state v_q. w_{v_q} is *within-constraint* if it can be obtained as the derivative of a smooth, within-constraint alteration of v_q. This occurs exactly if $w_{v_q} \in \boldsymbol{T}\mathcal{D}$ and w_{v_q} is the second derivative of a curve with derivative in $\mathcal{D}$, i.e. exactly if $w_{v_q} \in \mathcal{K}_{\mathcal{D}}$. $\mathcal{K}_{\mathcal{D}}$ is involutive by Lemma 2.1, and $\mathcal{K}_{\mathcal{D}}$ is $\boldsymbol{T}F_t$ invariant, since $Y_{\delta L} \in \mathcal{K}_{\mathcal{D}}$, so $F_t(w_{v_q}) \in \mathcal{K}_{\mathcal{D}}$. Thus $\delta q(t) = \boldsymbol{T}\tau_{\mathcal{Q}}\,\boldsymbol{T}F_t(w_{v_q}) \in \mathcal{D}$, the second (non-boundary) term of (11) vanishes, the first (boundary) term is successfully isolated, and the variational development of the symplecticity of the flow proceeds as in unconstrained Lagrangian systems.

If $\mathcal{D}$ is not involutive, corresponding to a nonholonomic system, then the second (non-boundary) term of (11) vanishes only under the condition that $\delta q(t) \in \mathcal{D}$.

This cannot be guaranteed, since, although $v_q \in \boldsymbol{T}\mathcal{D}$ is fully under control, $\delta q(t)$ for $t \neq 0$ is fully determined from that by the nonholonomic flow. *Nonholonomic systems exhibit nonsymplectic dynamics inasmuch as within-constraint perturbations yield out-of-constraint variations, because it is exactly this that obstructs the isolation of the boundary part of the Lagrange–d'Alembert principle.*

4.1. Symmetry and momentum

Suppose that a Lie group $\mathcal{G}$, with Lie algebra $\mathfrak{g}$, acts on $\mathcal{Q}$ such that $\mathcal{D}$ and L are invariant under the lifted action to $\boldsymbol{T}\mathcal{Q}$. Then the nonholonomic flow is equivariant. Choosing $\xi \in \mathfrak{g}$ and inserting the infinitesimal generator $w_{v_q} = \xi_{T\mathcal{Q}}(v_q)$ into (12) gives

$$\delta q(t) = \boldsymbol{T}\tau_{\mathcal{Q}}\,\boldsymbol{T}F_t\,\xi_{T\mathcal{Q}}(v_q) = \boldsymbol{T}\tau_{\mathcal{Q}}\,\xi_{T\mathcal{Q}}\,F_t(v_q) = \xi_{\mathcal{Q}}\big(q(t)\big),$$

so $\delta q \in \mathcal{D}$ if and only if $\xi_{\mathcal{Q}}(q) \in \mathcal{D}$ for all $q \in \mathcal{Q}$. Symmetries of this kind are called *vertical*. For a vertical symmetry ξ, the non-boundary term of (11) vanishes when $w_{v_q} = \xi_{T\mathcal{Q}}(v_q)$, the variational development is unobstructed, and conservation of the momentum J_ξ follows.

Fix $q \in \mathcal{Q}$, let $v_q \in \mathcal{D}$, and suppose $\xi \in \mathfrak{g}$ is such that $\xi_{\mathcal{Q}}(q) \in \mathcal{D}$, i.e. instead of assuming $\xi_{\mathcal{Q}}(q) \in \mathcal{D}$ for all q, just assume that at one fixed q. S_t is invariant because L is, and inserting $\xi_{T\mathcal{Q}}$ into (11) gives

$$\begin{aligned} 0 &= \boldsymbol{i}_{\xi_{T\mathcal{Q}}}\boldsymbol{d}S_t(v_q) \\ &= \boldsymbol{i}_{\xi_{T\mathcal{Q}}}(F_t^*\theta_L - \theta_L)(v_q) + \int_0^t \delta L(Y_{\delta L})\,\boldsymbol{T}_{v_q}(\tau_{\mathcal{Q}}\,F_s)\,\xi_{T\mathcal{Q}}(v_q)\,ds. \end{aligned}$$

Differentiate both sides of this equation with respect to t at $t = 0$. The integral term vanishes, since $\boldsymbol{T}\tau_{\mathcal{Q}}\,\xi_{T\mathcal{Q}}(v_q) = \xi_{\mathcal{Q}}\big(\tau_{\mathcal{Q}}(v_q)\big) \in \mathcal{D}$, so there results

$$\left.\frac{d}{dt}\right|_{t=0} \big(J_\xi\,F_t(v_q)\big) = \left.\frac{d}{dt}\right|_{t=0} F_t^*\big(\boldsymbol{i}_{\xi_{T\mathcal{Q}}}\theta_L\big)(v_q) = \left.\frac{d}{dt}\right|_{t=0} \boldsymbol{i}_{\xi_{T\mathcal{Q}}}\,F_t^*\theta_L(v_q) = 0.$$

Thus the time evolution of J_ξ has a critical point when it passes through a configuration q such that $\xi q \in \mathcal{D}$. If $\xi(t)$ is a curve in $\mathfrak{g}$ such that $\xi(t)q(t) \in \mathcal{D}$ for all t, then

$$\begin{aligned} \left.\frac{d}{dt}\right|_{t=t_0} J_{\xi(t)}\big(F_t(v_q)\big) &= \left.\frac{d}{dt}\right|_{t=t_0} J_{\xi(t_0)}\big(F_t(v_q)\big) + \left.\frac{d}{dt}\right|_{t=t_0} J_{\xi(t)}\big(F_{t_0}(v_q)\big) \\ &= \left.\frac{d}{dt}\right|_{t=t_0} J_{\xi(t)}\big(F_{t_0}(v_q)\big) \end{aligned}$$

which after replacing t_0 with t, is the momentum equation of [6]

$$\frac{d}{dt}\,J_{\xi(t)}\big(v(t)\big) = \left\langle J\big(v(t)\big), \frac{d\xi}{dt}(t)\right\rangle.$$

4.2. Symplecticity

Define the one-form α_t on $\mathcal{D}$ by

$$\alpha_t(w_{v_q}) \equiv \int_0^t \delta L\big(q''(s)\big) \cdot \delta q \, ds, \qquad q(t) = F_t(q), \qquad \delta q = \boldsymbol{T}(\tau_Q F_t) w_{v_q}.$$

This is the second (non-boundary) term of (11), which then becomes

$$\boldsymbol{d}S_t = F_t^* \theta_L - \theta_L + \alpha_t, \tag{13}$$

from which follows

$$F_t^* \omega_L - \omega_L = \boldsymbol{d}\alpha_t.$$

One sees that the flow is symplectic under the condition that the exterior derivative of the non-boundary term α_t vanishes, i.e. F_t is symplectic on pairs of vectors

$$\big\{(w_{v_q}, \tilde{w}_{v_q}) : w_{v_q}, \tilde{w}_{v_q} \in \boldsymbol{T}\mathcal{D} \text{ and } \boldsymbol{d}\alpha_t(w_{v_q}, \tilde{w}_{v_q}) = 0 \text{ for all } t\big\}.$$

This is the most general possible nonholonomic remnant of symplecticity.

Suppose one has a nonholonomic system and a holonomic system. The nonholonomic system has configuration space $\mathcal{Q}_1$, Lagrangian L_1, and nonholonomic constraint $\mathcal{D}_1$. This system is to be thought of as totally nonholonomic: it is not symmetric and has no symplectic remnant. The holonomic system has configuration space $\mathcal{Q}_2$, involutive constraint $\mathcal{D}_2$, and Lagrangian L_2. The direct product of these two systems is the nonholonomic system with configuration space $\mathcal{Q} = \mathcal{Q}_1 \times \mathcal{Q}_2$, nonholonomic constraint $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2$, and Lagrangian $L = L_1 + L_2$. This is a nonholonomic and holonomic system, with no interaction. A small coupling of the two systems can be introduced, obtaining a system having a component that is nearly holonomic. The issues are transparent here because of the explicit direct product, but a diffeomorphism of $\mathcal{Q}_1 \times \mathcal{Q}_2$ to another configuration space $\mathcal{Q}$ can destroy the direct product structure, whereas the physics would remain unaltered. The point is that one can easily arrange nonholonomic systems that have holonomic subsystems, as well as systems that are in some sense nearly holonomic.

For a direct product of a nonholonomic and holonomic system, such as above, it is natural to extract the following two subbundles, which in coordinates are

$$\mathcal{D}_0 \equiv \big\{(q_1, q_2, v_1, v_2) \in \mathcal{D}_1 \times \mathcal{D}_2 : v_1 = 0\big\},$$
$$\mathcal{D}_0' \equiv \big\{(q_1, q_2, v_1, v_2, \dot{q}_1, \dot{q}_2, \dot{v}_1, \dot{v}_2) \in \boldsymbol{T}\mathcal{D}_1 \times \boldsymbol{T}\mathcal{D}_2 : \dot{q}_1 = 0, \ \dot{v}_1 = 0\big\}.$$

The subbundle $\mathcal{D}_0$ contains the infinitesimal information about perturbing from one configuration to another holonomically. $\mathcal{D}_0'$, a subbundle of $\boldsymbol{T}\mathcal{D}$, contains the same information but for elements of phase space rather than configuration space. Note the following properties:

1. $\mathcal{D}_0 \subseteq \mathcal{D}$ and $[V, W] \in \mathcal{D}$ for all $V, W \in \mathcal{D}_0$;
2. $\boldsymbol{T}\tau_Q(D_0') \subseteq D_0$;
3. D_0' is invariant under the flow $\boldsymbol{T}F_t$, where F_t is the flow of $Y_{\delta L}$.

Call such a (D_0', D_0) a *symplecticity pair.* For the direct product system, a perturbation inside D_0' is a perturbation of only the second, holonomic system, so it is clear that

$$\omega_L\big(\boldsymbol{T}F_t(w_{v_q}), \boldsymbol{T}F_t(\tilde{w}_{v_q})\big) = \omega_L(w_{v_q}, \tilde{w}_{v_q})$$

for all $w_{v_q}, \tilde{w}_{v_q} \in D'_0$.

PROPOSITION 4.1. *If* (D'_0, D_0) *is a symplecticity pair then the (nonholonomic) flow* F_t *is symplectic when restricted to* D'_0.

Proof: It suffices to show that $\boldsymbol{d}\alpha_t(v'_q, w'_q) = 0$ for given $v'_q, w'_q \in \mathcal{D}'_0$. Find vector fields $V', W' \in \mathcal{D}'_0$ such that $V'(q) = v'_q$ and $W'(q) = w'_q$. Since $\boldsymbol{T}\tau_Q(\mathcal{D}'_0) \subseteq \mathcal{D}_0$, one can choose V', W' so there are vector fields $V, W \in \mathcal{D}_0$ such that

$$\boldsymbol{T}\tau_Q\, V' = V \circ \tau_Q, \qquad \boldsymbol{T}\tau_Q\, W' = W \circ \tau_Q. \tag{14}$$

Then

$$\boldsymbol{d}\alpha_t(V', W') = V'\big(\alpha_t(W')\big) - W'\big(\alpha_t(V')\big) - \alpha_t([V', W']). \tag{15}$$

The first term is zero since $\delta V'$ defined by $\delta V' = \boldsymbol{T}\tau_Q \boldsymbol{T}F_t\, V'$ is inserted as δq in definition of α_t and $\delta V'$ has values in $\mathcal{D}$ ($\boldsymbol{T}F_t$ maps $\mathcal{D}'_0$ to itself and $\boldsymbol{T}\tau_Q$ maps $\mathcal{D}'_0$ into $\mathcal{D}_0 \subseteq \mathcal{D}$). Similarly the second term of (15) vanishes, and so does the third term, because $\boldsymbol{T}\tau_Q[V', W'] = [V, W] \circ \tau_Q \in \mathcal{D}$ by (14) and the bracket of vector fields in $\mathcal{D}_0$ is in $\mathcal{D}$. □

4.3. Semi-Hamilton's equations

From Eq. (10), after replacing y_{v_q} with evaluations of $Y_{\delta L}$,

$$\boldsymbol{i}_{Y_{\delta L}}\omega_L(w_{v_q}) = \boldsymbol{d}E(w_{v_q}) + \delta L \circ Y_{\delta L}\big(\boldsymbol{T}\tau_Q w_{v_q}\big), \tag{16}$$

so, if $w_{v_q} \in \mathcal{K}_{\mathcal{D}}$, then

$$\boldsymbol{i}_{Y_{\delta L}}\omega_L(w_{v_q}) = \boldsymbol{d}E(w_{v_q}),$$

since $\delta L \circ Y_{\delta L}$ takes values in $\operatorname{ann}\mathcal{D}$, by definition of $Y_{\delta L}$. A key observation is that $Y_{\delta L}$ is determined by these equations, since, as will now be shown, the semi-symplectic form is nondegenerate on its subbundle. This, together with the computation just above, recovers Theorem 1 of [36]. Similar equations are found in [11, 16, 17, 31, 32].

THEOREM 4.1. *L is $\mathcal{D}$-regular if and only if ω_L is nondegenerate on $\mathcal{K}_{\mathcal{D}}$.*

Proof: In coordinates, the Euler–Lagrange form is

$$\omega_L = -\boldsymbol{d}\theta_L = \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}\, dq^i \wedge d\dot{q}^j + \frac{\partial^2 L}{\partial \dot{q}^i \partial q^j}\, dq^i \wedge dq^j.$$

As in Lemma 2.1, choosing basis vector fields X_α of $\mathcal{D}$, inserting the basis (3) gives a block matrix of the form

$$\begin{bmatrix} M_{11} & M_{12} \\ -M_{12} & 0 \end{bmatrix},$$

where M_{12} is the matrix of $\boldsymbol{F}^2L$ with respect to the basis X_α. So L is $\mathcal{D}$-regular if and only if M_{12} is nonsingular if and only if the block matrix has trivial kernel, as required. □

5. Semi-Hamiltonian systems

Given a distribution $\mathcal{D}$ and a $\mathcal{D}$-regular Lagrangian L, the two-form ω_L is nondegenerate on $\mathcal{K}_\mathcal{D}$ and the Lagrange–d'Alembert equations are the semi-Hamilton equations $\boldsymbol{i}_{Y_{\delta L}}\omega_L = \boldsymbol{d}E$ on $\mathcal{K}_\mathcal{D}$. If $\boldsymbol{F}L$ is a hyperregular (i.e. if L is diffeomorphism), then one can transfer this diffeomorphically to the cotangent bundle $\boldsymbol{T}^*\mathcal{Q}$, as follows: define $\mathcal{D}^* = \boldsymbol{F}L(\mathcal{D})$, use the canonical form $(\boldsymbol{F}L)_*\omega_L = \omega_0$, and the distribution $\mathcal{K}_\mathcal{D}$ becomes $\mathcal{K}^*_\mathcal{D} \equiv \boldsymbol{T}(\boldsymbol{F}L)\mathcal{K}_\mathcal{D} = \boldsymbol{T}\mathcal{D}^* \cap (\boldsymbol{T}\tau^*_\mathcal{Q})^{-1}\mathcal{D}$, as is easily checked. The phase space $\mathcal{D}^*$, distribution $\mathcal{K}^*_\mathcal{D}$, and ω_0 are the Hamiltonian formalism used in [4], and the relation between the two formalisms is considered by [18, 19]. The essentials of both formalisms are a phase space $\mathcal{P}$, the distribution of possible infinitesimal evolutions $\mathcal{K}$, and a nondegenerate two-form on $\mathcal{K}$. As in [35], where the tuple $(\mathcal{P}, \omega, \mathcal{K})$ is called a *distributional Hamiltonian system*, it is efficient and clarifying to make a formal definition.

DEFINITION 5.1. A *semi-symplectic manifold* is a tuple $(\mathcal{P}, \omega, \mathcal{K}, \mathcal{K}^\circ)$, where

1. $\mathcal{P}$ is a manifold;
2. $\mathcal{K}$ is a distribution on $\mathcal{P}$; $\mathcal{K}^\circ$ is a subbundle of $\mathcal{K}^*$;
3. ω is a 2-form defined on a distribution $\tilde{\mathcal{K}} \supseteq \mathcal{K}$ such that, for all $v_p \in \mathcal{K}$, $\omega(p)(v_p, w_p) = 0$ for all $w_p \in \mathcal{K}$ implies $v_p = 0$;

A *semi-symplectomorphism* $f : (\mathcal{P}_1, \omega_1, \mathcal{K}_1, \mathcal{K}^\circ_1) \to (\mathcal{P}_2, \omega_2, \mathcal{K}_2, \mathcal{K}^\circ_2)$ is a C^1 map $f : \mathcal{P}_1 \to \mathcal{P}_2$ such that

1. if $\alpha \in (\mathcal{K}^\circ_2)_{p_2}$ and $f(p_1) = p_2$ then $(T_{p_1}f)^*\alpha \in (\mathcal{K}^\circ_1)_{p_1}$;
2. $\omega_1 = f^*\omega_2$ on $\mathcal{K}_1$.

$\mathcal{K}^\bullet$ is the span of Y_f for all functions f such that $\boldsymbol{d}f \in \mathcal{K}$. $(\mathcal{P}, \omega, \mathcal{K})$ is *symplectic* if $\mathcal{K}^\circ = \mathcal{K}^*$, $\mathcal{K}$ is involutive, and $\boldsymbol{d}\omega = 0$.

As a convention, if in some context $\mathcal{K}^\circ$ is omitted (e.g. the semi-Hamiltonian system $(\mathcal{P}, \omega, \mathcal{K})$) then the context does not reference them and the statements are true for all choices of such subbundles. Since the definition of a semi-symplectomorphism references only various fibers over points, it makes sense to refer to a map as being semi-symplectic at a point or on a subset.

In passing from symplectic manifolds to semi-symplectic manifolds, ω can be evaluated in general only on vectors in $\mathcal{K}$, some nondegeneracy of ω is retained, and ω is not assumed to be closed. It is important that the definition allows that ω could be defined on a distribution $\tilde{\mathcal{K}}$ which strictly includes $\mathcal{K}$, and $\mathcal{K} = \boldsymbol{T}\mathcal{P}$ is a common special case. It is efficient to adopt the convention that unquantified variables range over whatever makes sense given the context. Especially, if the context is all semi-symplectic manifolds then ω can only be evaluated on elements of $\mathcal{K}$.

DEFINITION 5.2. Let $(\mathcal{P}, \omega, \mathcal{K}, \mathcal{K}^\circ)$ be a semi-symplectic manifold. The *semi-Hamiltonian vector field* of $f : P \to \mathbb{R}$ is the unique vector field Y_f such that $Y_f(p) \in \mathcal{K}$ for all $p \in P$ and $\boldsymbol{i}_{Y_f}\omega(v) = \boldsymbol{d}f(v)$ for all $v \in \mathcal{K}$. If $f, g : P \to \mathbb{R}$ then the *almost-Poisson bracket* of $f, g : \mathcal{P} \to \mathbb{R}$ is $\{f, g\} \equiv \omega(Y_f, Y_g)$.

DEFINITION 5.3. A *semi-Hamiltonian system* is a tuple $(\mathcal{P}, \omega, \mathcal{K}, \mathcal{K}^\circ, H)$ such that $\mathcal{P}$ is a semi-symplectic manifold and $H : \mathcal{P} \to \mathbb{R}$ satisfies $\boldsymbol{d}H \in \mathcal{K}^\circ$. A *time-evolution* is an integral curve of Y_H. $(\mathcal{P}, \omega, \mathcal{K}, \dot{\mathcal{K}}, \mathcal{K}^\circ, H)$ is *Hamiltonian* if $(\mathcal{P}, \omega, \mathcal{K}, \mathcal{K}^\circ)$ is symplectic.

The categories of symplectic manifolds and Hamiltonian systems are recovered in the special case that $\mathcal{K}^\circ = \boldsymbol{T}^*\mathcal{P}$.

The following basic properties follow easily.

PROPOSITION 5.1. *Let $(\mathcal{P}, \omega, \mathcal{K})$ be a semi-symplectic manifold. Then*

1. $Y_{af+bg} = aY_f + bY_g$, $Y_{fg} = fY_g + gY_f$.
2. $\{af+bg, h\} = a\{f, h\}+b\{g, h\}$, $\{f, g\} = -\{g, f\}$, $\{fg, h\} = f\{g, h\}+g\{f, h\}$.
3. *If ψ is C^1 and semi-symplectic, then $\boldsymbol{T}\psi \circ Y_{f\circ\psi} = Y_f \circ \psi$.*

The almost-Poisson bracket satisfies all the properties of a Poisson bracket except the Jacobi identity. It makes little sense to call the semi-symplectic category "almost symplectic" since many of the structural aspects are broken in the generalization: "half symplectic" is more appropriate. Almost-Poisson brackets are covered by existing standard theory, as in [27]. The principle fact is that there is a $\binom{2}{0}$ *Poisson tensor* B such that $\{f, g\} = B(\boldsymbol{d}f, \boldsymbol{d}g)$. The Hamiltonian vector field is $Y_f = B^\flat \boldsymbol{d}f$, so that $\mathcal{K} = \text{image}\, B^\flat$ and so the characteristic field of B is the distribution $\mathcal{K}$. Clearly, the category of semi-symplectic manifolds, with $\mathcal{K}^\circ = \mathcal{K}^*$, is equivalent to the category of almost-Poisson manifolds with characteristic field that is a distribution.

Since $Y_f \in \mathcal{K}$,

$$\boldsymbol{i}_{Y_f}\boldsymbol{d}g = \omega(Y_g, Y_f) = \{g, f\}$$

(the first equality is false in general if Y_f is replaced by a vector field X that is not in $\mathcal{K}$). Consequently

$$\frac{df}{dt} = \boldsymbol{i}_{Y_H}\boldsymbol{d}f = \{f, H\},$$

so the energy H is conserved since the almost-Poisson bracket is antisymmetric. The almost-Poisson bracket and the almost-Poisson equations of motion occur exactly as above in [35], and somewhat earlier in the context of cotangent bundles in [2] and [23].

Some important structural aspects of symplectic manifolds (Hamiltonian systems) are broken in the generalization to semi-symplectic manifolds (semi-Hamiltonian systems). This is caused mainly by two relaxations:

1. $\boldsymbol{i}_{Y_f}\omega = \boldsymbol{d}f$ is necessarily true only on $\mathcal{K}$; and
2. ω is only defined on $\mathcal{K}$.

For example, $\boldsymbol{d}(\boldsymbol{i}_{Y_f}\omega) = 0$ fails in the context of semi-symplectic manifolds because $\boldsymbol{i}_{Y_f}\omega = \boldsymbol{d}f$ fails. The flow F_t of Y_f need not preserve $\mathcal{K}$, i.e. $\boldsymbol{T}F_t(\mathcal{K}) \subseteq \mathcal{K}$ fails, so $\boldsymbol{L}_{Y_f}\omega = 0$ cannot be defined in general, because ω is only defined on $\mathcal{K}$ in general. Even if ω is defined on all of $\boldsymbol{T}\mathcal{P}$ and $\boldsymbol{d}\omega = 0$, $\boldsymbol{L}_{Y_f}\omega = \boldsymbol{d}(\boldsymbol{i}_{Y_f}\omega) + \boldsymbol{i}_{Y_f}\boldsymbol{d}\omega = \boldsymbol{d}(\boldsymbol{i}_{Y_f}\omega)$ which is not zero, so the flow of a semi-Hamiltonian vector field is not symplectic. The Lie bracket $[Y_f, Y_g]$ is not in $\mathcal{K}$ so it cannot equal $Y_{\{f,g\}}$, because that always is in $\mathcal{K}$. Thus the semi-Hamiltonian vector fields do not generally form a Lie algebra.

Since the case of ω nondegenerate is common, a few deeper structural properties of the almost-Poisson bracket are worth stating.

PROPOSITION 5.2. *Let $(\mathcal{P}, B)$ be an almost-Poisson manifold. Then*

1. $-\big([Y_f, Y_g] + Y_{\{f,g\}}\big)(h) = \{\{f, g\}, h\} + \circlearrowleft$.
2. $\boldsymbol{L}_{Y_f} B(\boldsymbol{d}g, \boldsymbol{d}h) = \frac{1}{2}\boldsymbol{i}_{[B,B]}\boldsymbol{d}f \wedge \boldsymbol{d}g \wedge \boldsymbol{d}h = \{\{f, g\}, h\} + \circlearrowleft$.

Proof: For the first item,

$$\begin{aligned} -\big([Y_f, Y_g] + Y_{\{f,g\}}\big)(h) &= -Y_f\big(Y_g(h)\big) + Y_g\big(Y_f(h)\big) - Y_{\{f,g\}}(h) \\ &= -\{\{h, g\}, f\} + \{\{h, f\}, g\} - \{h, \{f, g\}\} \\ &= \{\{f, g\}, h\} + \circlearrowleft . \end{aligned}$$

The second item is standard for almost-Poisson manifolds and is found in [27]. □

The semi-symplectic fact corresponding to the second item of Proposition 5.2 requires that $\mathcal{K}$ be involutive to well define $\boldsymbol{L}$ and $\boldsymbol{d}$.

PROPOSITION 5.3. *If $(P, \mathcal{K}, \omega)$ be a semi-symplectic manifold and $\mathcal{K}$ is involutive, then*

$$-\frac{1}{2}\boldsymbol{L}_{Y_f}\omega(Y_g, Y_h) = -\frac{1}{2}\boldsymbol{d}\omega(Y_f, Y_g, Y_h) = \{\{f, g\}, h\} + \circlearrowleft .$$

Proof: Using (16) of Theorem 5.1 with $\mathcal{K}_0 = \mathcal{K}$,

$$\boldsymbol{L}_{Y_f}\omega(Y_g, Y_h) = (\boldsymbol{i}_{Y_f}\boldsymbol{d}\omega)(Y_g, Y_h)$$

so the first equals the second. The second equals the third because

$$\begin{aligned} \boldsymbol{d}\omega(Y_f, Y_g, Y_h) &= \Big(Y_f\big(\omega(Y_g, Y_h)\big) - \omega([Y_f, Y_g], Y_h)\Big) + \circlearrowleft \\ &= -2(\{\{f, g\}, h\} + \circlearrowleft). \end{aligned}$$ □

Since the categories of semi-Poisson (with characteristic field a distribution) and semi-Hamiltonian manifolds are equivalent, the conditions under which they specialize to the usual Poisson and symplectic geometry are equivalent This is demonstrated in Proposition 5.4, a cotangent-bundle coordinate version of which is found in [2, 38].

PROPOSITION 5.4. $[B, B] = 0$ *if and only if $\mathcal{K}$ is involutive and $\boldsymbol{d}\omega = 0$, i.e. (P, B) is Poisson if and only if $(\mathcal{P}, \mathcal{K}, \omega)$ is symplectic.*

Proof: If $[B, B] = 0$ then $[Y_f, Y_g] = -Y_{\{f,g\}}$ and $\mathcal{K}$ is locally spanned by semi-Hamiltonian vector fields so $\mathcal{K}$ is integrable. Also, $[B, B] = 0$ implies $\{\{f, g\}, h\} + \circlearrowleft = 0$ so $d\omega(Y_f, Y_g, Y_h) = 0$ which implies $d\omega = 0$. Conversely, if $\mathcal{K}$ is involutive and $d\omega = 0$ then $\{\{f, g\}, h\} + \circlearrowleft = 0$ for all f, g, h so $[B, B] = 0$. □

Under some conditions the semi-Hamiltonian time evolution shows remnants of the symplectic category. Theorem 5.1 shows that if $\mathcal{K}$ admits an ideal $\mathcal{K}_0$ on which $d\omega = 0$ then the flow of any semi-Hamiltonian vector field is symplectic on $\mathcal{K}_0$.

THEOREM 5.1. *Suppose $(\mathcal{P}, \omega, \mathcal{K})$ is a semi-symplectic manifold and that $\mathcal{K}_0$ is a subbundle of $\mathcal{K}$ such that $[X, Y] \in \mathcal{K}_0$ for all $X \in \mathcal{K}$ and $Y \in \mathcal{K}_0$. Then, for all flows F_t of semi-Hamiltonian vector fields Y_f,*

1. *$\boldsymbol{T}F_t(\mathcal{K}_0) \subseteq \mathcal{K}_0$ (also true for the flow of any vector field in $\mathcal{K}$); and*
2. *$\boldsymbol{L}_{Y_f}\omega = \boldsymbol{i}_{Y_f}\boldsymbol{d}\omega$ on $\mathcal{K}_0$.*

Proof: The first item follows from $[X, Y] \in \mathcal{K}_0$ for all $Y \in \mathcal{K}_0$ and Lemma 2.2. For the second item, if $Y_1, Y_2 \in \mathcal{K}_0$ then $[Y_1, Y_2] \in \mathcal{K}_0$ and

$$\begin{aligned} \boldsymbol{d}(\boldsymbol{i}_{Y_f}\omega)(Y_1, Y_2) &= Y_1\big(\boldsymbol{i}_{Y_f}\omega(Y_2)\big) - Y_2\big(\boldsymbol{i}_{Y_f}\omega(Y_1)\big) - \boldsymbol{i}_{Y_f}\omega([Y_1, Y_2]) \\ &= Y_1\big(Y_2(f)\big) - Y_2\big(Y_1(f)\big) - [Y_1, Y_2](f) \\ &= 0, \end{aligned}$$

so that

$$\boldsymbol{L}_{Y_f}\omega(Y_1, Y_2) = \boldsymbol{d}(\boldsymbol{i}_{Y_f}\omega)(Y_1, Y_2) + \boldsymbol{i}_{Y_f}\boldsymbol{d}\omega(Y_1, Y_2) = \boldsymbol{i}_{Y_f}\boldsymbol{d}\omega(Y_1, Y_2), \tag{17}$$

as required. □

Theorem 5.1 is cast to make purely Lie theoretic assumptions on the vector fields in $\mathcal{K}_0$. Essentially the same proof gives the following less coarse result. The motivation is there might be invariant sets, such as equilibria, or relative equilibria, where symplectic effects emerge that do not occur in the whole phase space.

THEOREM 5.2. *Let $(\mathcal{P}, \omega, \mathcal{K}, H)$ be a semi-Hamiltonian system, and let $\mathcal{K}_0$ be a subbundle of $\mathcal{K}$ over $\mathcal{P}_0 \subseteq \mathcal{P}$. Suppose that*

1. *$\mathcal{K}_0$ is $\boldsymbol{T}F_t$ invariant, where F_t is the flow of Y_H;*
2. *$\Delta_{\mathcal{K}} = 0$ on $\mathcal{K}_0$;*
3. *$\boldsymbol{d}\omega = 0$ on $\mathcal{K}_0$.*

Then F_t is symplectic on $\mathcal{K}_0$.

Proof: The assumption $\Delta_{\mathcal{K}} = 0$ on $\mathcal{K}_0$ implies $[X, Y] \in \mathcal{K}$ for all X, Y with values (on $\mathcal{P}_0$) in $\mathcal{K}_0$. Then the computation in Theorem 5.1 of $\boldsymbol{d}(\boldsymbol{i}_{Y_f}\omega)$ and $\boldsymbol{L}_{Y_H}\omega$, with f replaced by H, gives in the same way $\boldsymbol{L}_{Y_H}\omega = 0$ and hence the result. □

The relationship between this theorem and the symplecticity criteria from the variational development of Section 4.2 is just that one can take $\mathcal{K}_0$ to be the subbundle $\mathcal{D}'_0$ of Proposition 4.1.

If $\mathcal{K}$ is not involutive then generally there are not integral submanifolds through every point. However, this does not preclude the presence of submanifolds with tangent bundle in $\mathcal{K}$; for example, there are always one-dimensional submanifolds with this property by taking the integral curves of any vector field in $\mathcal{K}$. From the nonholonomic point of view, a submanifold $\mathcal{M} \subseteq \mathcal{P}$ such that $\boldsymbol{T}\mathcal{M} \subseteq \mathcal{K}$ corresponds to a subset of phase space within which every curve is consistent with the constraint.

COROLLARY 5.1. *Let $(\mathcal{P}, \omega, \mathcal{K}, H)$ be a semi-Hamiltonian system, and let $\mathcal{M} \subseteq \mathcal{P}$ be a submanifold of $\mathcal{P}$ such that $\boldsymbol{T}\mathcal{M} \subseteq \mathcal{K}$ and such that $\mathcal{M}$ is invariant under the flow F_t of Y_H. Suppose the pull back of ω to $\mathcal{M}$ is closed. Then F_t is symplectic on $\mathcal{M}$.*

Proof: $\boldsymbol{T}\mathcal{M}$ is invariant under $\boldsymbol{T}F_t$ because $\mathcal{M}$ in invariant under F_t. The vector fields tangent to $\mathcal{M}$ are closed under Lie bracket and therefore $\Delta_{\mathcal{K}} = 0$ on $\boldsymbol{T}\mathcal{M}$. $\boldsymbol{d}\omega = 0$ on $\boldsymbol{T}M$ is equivalent to the assumption that the pull back of ω to $\mathcal{M}$ is closed. So, taking $\mathcal{K}_0 = \boldsymbol{T}\mathcal{M}$, Corollary follows directly from Theorem 5.2. □

The semi-symplectic systems arising from the Lagrange–d'Alembert principle have the property that the semi-symplectic form is the restriction of a symplectic form defined on the whole phase space. This special case is the most common, and has some additional features that deserve mention.

Suppose $(\mathcal{P}, \omega, \mathcal{K})$ is a semi-symplectic manifold and ω is defined on $\boldsymbol{T}\mathcal{P}$. For $w \in \boldsymbol{T}_p\mathcal{P}$ define $\langle \alpha_w, v\rangle = \omega(w, v)$. Then there is a unique $\kappa(w) \in \mathcal{K}_p$ such that $\omega\big(\kappa(w), z\big) = \alpha_v(z)$ for all $z \in \mathcal{K}_p$, so $\kappa : \boldsymbol{T}\mathcal{P} \to \mathcal{K}$. Since $\kappa^2 = \kappa$ and $\operatorname{image}\kappa = \mathcal{K}$, κ is a projection onto $\mathcal{K}$. Since $\ker\kappa = \mathcal{K}^{\omega\perp}$, $\boldsymbol{T}\mathcal{P} = \operatorname{image}\kappa \oplus \ker\kappa = \mathcal{K} \oplus \mathcal{K}^{\omega\perp}$.

With ω defined on $\boldsymbol{T}\mathcal{P}$ and not just on $\mathcal{K}$, constructs such as $\boldsymbol{L}_{Y_f}\omega$ are well defined even when $\mathcal{K}$ is not involutive. This and the extra property that $\mathcal{K}$ has a natural complement, gives a variety of facts:

1. $\boldsymbol{i}_{Y_f}\omega = \kappa^*\boldsymbol{d}f = \boldsymbol{d}_\kappa f$.
2. $\boldsymbol{L}_{Y_f}\omega = \boldsymbol{d}(\boldsymbol{d}_\kappa f)$, so $\kappa^*\boldsymbol{L}_{Y_f}\omega = \boldsymbol{d}_\kappa^2 f = \langle \boldsymbol{d}f, \Delta_\kappa\rangle$.
3. $\boldsymbol{i}_{[Y_f, Y_g]\omega} + \boldsymbol{d}\{f, g\} = \boldsymbol{i}_{Y_f}\langle \boldsymbol{d}g, \Delta_\kappa\rangle - \boldsymbol{i}_{Y_g}\langle \boldsymbol{d}f, \Delta_\kappa\rangle$: Since ω is closed,

$$\begin{aligned}\boldsymbol{i}_{[X,Y]}\omega &= \boldsymbol{L}_X\boldsymbol{i}_Y\omega - \boldsymbol{i}_Y\boldsymbol{L}_X\omega\\ &= \boldsymbol{d}(\boldsymbol{i}_X\boldsymbol{i}_Y\omega) + \boldsymbol{i}_X\boldsymbol{d}(\boldsymbol{i}_Y\omega) - \boldsymbol{i}_Y\big(\boldsymbol{d}(\boldsymbol{i}_X\omega) + \boldsymbol{i}_X\boldsymbol{d}\omega\big)\\ &= \boldsymbol{d}(\boldsymbol{i}_X\boldsymbol{i}_Y\omega) + \boldsymbol{i}_X\boldsymbol{d}(\boldsymbol{i}_Y\omega) - \boldsymbol{i}_Y\boldsymbol{d}(\boldsymbol{i}_X\omega),\end{aligned}$$

(this is true for any closed form) so that

$$\begin{aligned}\boldsymbol{i}_{[Y_f, Y_g]}\omega + \boldsymbol{d}\{f, g\} &= \boldsymbol{i}_{[Y_f, Y_g]}\omega - \boldsymbol{d}(\boldsymbol{i}_{Y_f}\boldsymbol{i}_{Y_g}\omega)\\ &= \boldsymbol{i}_{Y_f}\boldsymbol{d}(\boldsymbol{i}_{Y_g}\omega) - \boldsymbol{i}_{Y_g}\boldsymbol{d}(\boldsymbol{i}_{Y_f}\omega)\\ &= \boldsymbol{i}_{Y_f}\boldsymbol{d}_\kappa^2 g - \boldsymbol{i}_{Y_g}\boldsymbol{d}_\kappa^2 f\\ &= \boldsymbol{i}_{Y_f}\langle \boldsymbol{d}g, \Delta_\kappa\rangle - \boldsymbol{i}_{Y_g}\langle \boldsymbol{d}f, \Delta_\kappa\rangle.\end{aligned}$$

4. $\{\{f,g\},h\}+\circlearrowleft = i_{Y_f} i_{Y_g}\langle dh, \Delta_\kappa\rangle + \circlearrowleft$: Since $d\omega = 0$,

$$\begin{aligned} 0 = d\omega(Y_f, Y_g, Y_h) &= \Big(Y_f\big(\omega(Y_g, Y_h)\big) - \omega\big([Y_f, Y_g], Y_h\big)\Big) + \circlearrowleft \\ &= \Big(\{\{g,h\}, f\} - i_{Y_h}\big(i_{Y_f}\langle dg, \Delta_\kappa\rangle \\ &\qquad - i_{Y_g}\langle df, \Delta_\kappa\rangle - d\{f,g\}\big)\Big) + \circlearrowleft \\ &= \Big(2\{\{f,g\},h\} - 2 i_{Y_f} i_{Y_g}\langle dh, \Delta_\kappa\rangle\Big) + \circlearrowleft . \end{aligned}$$

5. A vector field $X \in \mathcal{K}$ is locally semi-Hamiltonian if $L_X\omega = 0$: $L_X\omega = i_X d\omega + d(i_X\omega) = d(i_X\omega)$ so $L_X\omega = 0$ implies $d(i_X\omega) = 0$, i.e. $i_X\omega = df$ locally.

In the case that $\mathcal{P}$ is a submanifold of a manifold $\hat{\mathcal{P}}$, and ω is the pull-back of a symplectic form $\hat{\omega}$, there is a natural projection onto $\mathcal{K}$ from the splitting $T\hat{\mathcal{P}} = \mathcal{K} \oplus \mathcal{K}^{\hat{\omega}\perp}$ and Y_H is the projection of X_H. This or similar projections are sometimes used to define Y_H in [10, 13, 15, 22, 23, 34]. In a Lagrangian system the definition of regularity can be cast in terms of such splittings: for example, by Theorem 4.1 the splitting $T\hat{\mathcal{P}} = \mathcal{K} \oplus \mathcal{K}^{\hat{\omega}\perp}$ occurs if and only if L is regular in a constrained sense.

5.1. Vector fields, symmetry, momentum

In mechanics, vector fields 1) generate symmetries, and 2) correspond to momenta. Here "momentum f corresponding to a vector field X" means X and f are related by $X = Y_f$. For a Hamiltonian system the two classes of vector fields are, respectively, 1) those X such that $L_X\omega = 0$, and 2) those X such that $d(i_X\omega) = 0$. These two classes are the same because $d\omega = 0$ implies $L_X\omega = d(i_X\omega)$. The two classes of vector fields are not the same for semi-Hamiltonian systems, because of two additional distinct conditions: 1) the infinitesimal condition that corresponds to preservation of $\mathcal{K}$ is $[X, Y] \in \mathcal{K}$ for all $Y \in \mathcal{K}$, whereas 2) every vector field of the form Y_f is in $\mathcal{K}$. The distinctness is one of "preservation" as opposed to "membership", and it results in vector fields that generate symmetries but do not correspond to momenta, and vector fields that correspond to momenta but do not generate symmetries.

DEFINITION 5.4. Let $(\mathcal{P}, \mathcal{K}, \omega, \mathcal{K}^\circ)$ be semi-symplectic and let X be a vector field on $\mathcal{P}$.

1. X is *locally semi-Hamiltonian* if for all p there is a function f defined on an open neighborhood of $U \ni p$ such that $X = Y_f$ on U.
2. X is *infinitesimally semi-symplectic* if
   (a) $L_X\omega = 0$;
   (b) $[X, Y] \in \mathcal{K}$ for all $Y \in \mathcal{K}$;
   (c) $L_X\alpha \in \mathcal{K}^\circ$ for all $X \in \mathcal{K}$ and $\alpha \in \mathcal{K}^\circ$.

3. X is *Noether* if X is locally semi-Hamiltonian and X is infinitesimally semi-symplectic.

If X is locally semi-Hamiltonian then $X \in \mathcal{K}$ since every semi-Hamiltonian vector field is in $\mathcal{K}$. If H is a Hamiltonian such that $X(H) = 0$ and $X = Y_f$ then f is conserved because $\dot{f} = \{f, H\} = -\{H, f\} = -Y_f(H) = 0$. Thus the semi-Hamiltonian vector fields correspond to the vector fields that generate conserved momenta. The flow F_t of an infinitesimally semi-symplectic vector field is a semi--symplectomorphism as follows easily from Lemma 2.2. Noether vector fields generate symmetries and also correspond to momenta.

If $\mathcal{K}$ is involutive, $\mathcal{K}^\circ = \mathcal{K}$, and $\boldsymbol{d}\omega = 0$, i.e. if $(\mathcal{P}, \mathcal{K}, \omega, \mathcal{K}^\circ)$ is symplectic, then $X \in \mathcal{K}$ is locally semi-Hamiltonian if and only if $\boldsymbol{L}_X\omega = 0$, in which case it is infinitesimally semi-symplectic, and hence Noether. Semi-symplectic transformations do not necessarily preserve the leaves of $\mathcal{K}$ (they might transform one leaf symplectically to another), so infinitesimally semi-symplectic vectors fields need not be in $\mathcal{K}$, and hence need not be Noether, even if $(\mathcal{P}, \mathcal{K}, \omega)$ is symplectic.

The different classes of infinitesimal symmetry foreshadow different classes of symmetry. Given an action of a Lie group on $\mathcal{P}$, some of it may result in genuine symmetry, some of it may result in conserved quantities, and some of it might be irrelevant.

DEFINITION 5.5. Let $(\mathcal{P}, \omega, \mathcal{K}, \mathcal{K}^\circ)$ be a semi-symplectic manifold and suppose Φ is a C^1 action of a Lie group $\mathcal{G}$ on $\mathcal{P}$.

1. $g \in \mathcal{G}$ is *semi-symplectic* if Φ_g is semi-symplectic. The *semi-symplectic subgroup* is the subgroup

$$\mathcal{G}^{\mathrm{ss}} \equiv \{g \in \mathcal{G} : g \text{ is semi-symplectic}\}.$$

ξ is *semi-symplectic* if it is in the Lie algebra $\mathfrak{g}^{\mathrm{ss}}$ of $\mathcal{G}^{\mathrm{ss}}$.

2. $\xi \in \mathfrak{g}$ is *semi-Hamiltonian* if $\xi_{\mathcal{P}}$ is locally semi-Hamiltonian. The *semi--Hamiltonian subspace* is $\mathfrak{g}^{\mathrm{sH}} \equiv \{\xi \in \mathfrak{g} : \xi \text{ is semi-Hamiltonian}\}$.
3. $\xi \in \mathfrak{g}$ is *Noether* if it is semi-Hamiltonian and semi-symplectic. The *Noether subspace* is $\mathfrak{g}^{\mathfrak{N}} \equiv \mathfrak{g}^{\mathrm{ss}} \cap \mathfrak{g}^{\mathrm{sH}}$.

All of these definitions can be localized to a point or subset, e.g. g is semi-symplectic at $p \in \mathcal{P}$ if Φ_g is semi-symplectic at p.

If g is semi-symplectic and ξ is semi-symplectic [semi-Hamiltonian] [Noether] then $\mathrm{Ad}_g\,\xi$ is semi-symplectic [semi-Hamiltonian] [Noether], respectively: Given a semi-symplectic g, $\mathrm{Ad}_g\,\xi$ is semi-symplectic because the Lie algebra of $\mathcal{G}^{\mathrm{ss}}$ is invariant under the adjoint action on any of its elements. If ξ is semi-Hamiltonian, $p \in \mathcal{P}$, and $\xi_{\mathcal{P}} = Y_f$ near p then

$$(\mathrm{Ad}_g\,\xi)_{\mathcal{P}} = (\Phi_g)_*\xi_{\mathcal{P}} = Y_{(\Phi_g)_*f},$$

so $\mathrm{Ad}_g\,\xi$ is also semi-Hamiltonian. If $g \in \mathcal{G}^{\mathrm{ss}}$, $\xi \in \mathfrak{g}^{\mathfrak{N}}$, and $p \in \mathcal{P}$, then $\xi(g^{-1}p) \in \mathcal{K}$,

$$(g\xi)p = g\big(\xi(g^{-1}p)\big) \in \mathcal{K},$$

since $\mathcal{K}$ is invariant. It follows that the Noether subspace is an ideal of $\mathfrak{g}^{\mathrm{ss}}$. The Lie algebra $\mathfrak{g}^{\mathrm{ss}}$ of $\mathcal{G}^{\mathrm{ss}}$ is the set of $\xi \in \mathfrak{g}$ such that $[\xi_{\mathcal{P}}, X] \in \mathcal{K}$ for all vector fields $X \in \mathcal{K}$, $\boldsymbol{L}_{\xi_{\mathcal{P}}}\alpha \in \mathcal{K}^\circ$ for all one-forms $\alpha \in \mathcal{K}^\circ$, and $\boldsymbol{L}_{\xi_{\mathcal{P}}}\omega = 0$.

DEFINITION 5.6. Let $(\mathcal{P}, \mathcal{K}, \omega)$ be a semi-symplectic manifold. $J : \mathcal{P} \to \mathfrak{g}^*$ is a *momentum mapping* for an action of a Lie group $\mathcal{G}$ on $\mathcal{P}$ if, for all $(p, \xi) \in \mathcal{P} \times \mathfrak{g}$ such that $\xi p \in \mathcal{K}$, $\boldsymbol{d}J_\xi(v_p) = (\boldsymbol{i}_{\xi_{\mathcal{P}}}\omega)(v_p)$.

The definition of a momentum mapping deliberately avoids quantifying $v_p \in \mathcal{K}_p$. This is important for conservation of momentum in the presence of nonlinear constraints (Section 6). In the examples the momentum mapping satisfies $\boldsymbol{d}J_\xi(v_p) = (\boldsymbol{i}_{\xi_{\mathcal{P}}}\omega)(v_p)$ for all v_p in the distribution on which ω is defined.

THEOREM 5.3. *If $\xi \in \mathfrak{g}$ is such that $\xi_{\mathcal{P}}(p) \in \mathcal{K}$ for all p, and if H is invariant, then ξ is semi-Hamiltonian and J_ξ is conserved. If $p(t) \in \mathcal{P}$ and $\xi(t) \in \mathfrak{g}$ are such that $p(t)$ is an evolution and $\xi(t)p(t) \in \mathcal{K}$ for all t, then*

$$\frac{d}{dt} J_{\xi(t)}\big(p(t)\big) = \left\langle J\big(p(t)\big), \frac{d\xi}{dt}(t)\right\rangle.$$

Proof: If $\xi_{\mathcal{P}}(p) \in \mathcal{K}$ for all p then $\xi_{\mathcal{P}}$ is, by the definition of the momentum map, the semi-Hamiltonian vector field corresponding to J_ξ. Thus $\xi_{\mathcal{P}}$ is locally semi-Hamiltonian so ξ is semi-Hamiltonian, and

$$\frac{dJ_\xi}{dt} = J_\xi(Y_H) = \omega(\xi_{\mathcal{P}}, Y_H) = -\omega(Y_H, \xi_{\mathcal{P}}) = \boldsymbol{d}H(\xi_{\mathcal{P}}) = 0.$$

If $\big(\xi(t)\big)_{\mathcal{P}}\big(p(t)\big) \in \mathcal{K}$ then

$$\begin{aligned}\left.\frac{d}{dt}\right|_{t=t_0} J_{\xi(t)}\big(p(t)\big) &= \left.\frac{d}{dt}\right|_{t=t_0} J_{\xi(t_0)}\big(p(t)\big) + \left.\frac{d}{dt}\right|_{t=t_0} J_{\xi(t)}\big(p(t_0)\big)\\ &= \left\langle J\big(p(t_0)\big), \frac{d\xi}{dt}(t_0)\right\rangle\end{aligned}$$

(the first summand of the middle equality is zero as in the proof of conservation of momentum). □

Symmetric semi-Hamiltonian systems admit a structural possibility which does not occur in Hamiltonian systems: there might be functions the semi-Hamiltonian vector fields of which are not infinitesimal generators but take values in $\mathfrak{g} \cap \mathcal{K}$. Such a function j is conserved by the flow of the semi-Hamiltonian vector field of any invariant Hamiltonian because

$$\frac{dj}{dt} = \boldsymbol{i}_{Y_H}\boldsymbol{d}j = -\boldsymbol{i}_{Y_j}\boldsymbol{d}H = 0.$$

However, unlike in the Hamiltonian case, j is independent of the conserved momenta because not all Lie algebra elements generate momenta. In this way symmetry-related conserved quantities occur which have been *orphaned* due to a condition where symmetry associated momenta have been *exiled*. To keep track of this situation, for a nonholonomic system with symmetry, one can choose $\mathcal{K}^\circ \supseteq \operatorname{ann}\mathfrak{g} \cap \mathcal{K}$.

5.2. Reduction

Whether or not the category of a semi-symplectic manifolds is central for nonholonomic mechanics depends in part on whether or not it is closed under appropriate notions of quotient, i.e. reductions of semi-Hamiltonian systems should again be semi-Hamiltonian. In the presence of an *a priori* symmetry and an invariant Hamiltonian H, appropriate means the following:

1. The reduced phase space should be constructed without using information specific to H; it should be the same for all invariant H.
2. The reduction should be optimal (c.f. [29]); the reduced space should be generic in the prevailing category. It should have no unaccounted for symmetry-related invariant sets or any residual symmetry, which would be present for all invariant H.

The development below follows [4]. For reduction theory using the Poisson formalism, see [35].

Let $(\mathcal{P}, \omega, \mathcal{K}, \mathcal{K}^\circ)$ be a semi-symplectic system, and let a Lie group $\mathcal{G}$ act on $\mathcal{P}$. Assume that $J : \mathcal{P} \to \mathfrak{g}^*$ is an equivariant momentum map.

DEFINITION 5.7. A *reduction datum* is a tuple $(\mathcal{H}, \mathfrak{n}, \mu)$ where $\mathcal{H}$ is a closed subgroup of $\mathcal{G}$, $\mathfrak{n}$ is a subspace of $\mathfrak{g}$, and $\mu \in \mathfrak{n}^*$, such that:

1. every $\xi \in \mathfrak{n}$ is *semi-Hamiltonian* on $(\iota_\mathfrak{n}^* J)^{-1}(\mu)$, where $\iota_\mathfrak{n} : \mathfrak{n} \to \mathfrak{g}$ is the injection;
2. every $h \in \mathcal{H}$ is semi-symplectic on $(\iota_\mathfrak{n}^* J)^{-1}(\mu)$;
3. $\mathfrak{n}$ is invariant under the adjoint action of $\mathcal{H}$;
4. $h\mu = \mu$ for all $h \in \mathcal{H}$.

$(\mathcal{H}, \mathfrak{n}, \mu)$ *contains* $p \in \mathcal{P}$ if $p \in (\iota_\mathfrak{n}^* J)^{-1}(\mu)$.

For example, one can take $\mathcal{H} \equiv \mathcal{G}^{\mathrm{ss}}$ and $\mathfrak{n} \equiv \mathfrak{g}^{\mathrm{sH}}$. For any reduction datum $(\mathcal{H}, \mathfrak{n}, \mu)$, $\mathfrak{n} \cap \mathfrak{h}$ is an ideal and $\mathcal{H}$ acts on $\mathfrak{n}^*$ by the coadjoint action. The definition of a reduction datum is motivated by the observation that the reduction procedure does not require global conditions such as $\xi p \in \mathcal{K}$ for all $p \in \mathcal{P}$. Rather, it is sufficient to have such conditions on an *a priori* level set of momenta, potentially leading to sharper reduced spaces. Failure to localize like this could result in reduced spaces with hidden residual symmetries or hidden conserved quantities.

Reduction data are partially ordered by

$$(\mathcal{H}, \mathfrak{n}, \mu) \leq (\mathcal{H}', \mathfrak{n}', \mu') \quad \Leftrightarrow \quad \mathcal{H} \subseteq \mathcal{H}', \qquad \mathfrak{n} \subseteq \mathfrak{n}', \qquad \mu = \mu'|\mathfrak{n}.$$

Fix $p_0 \in \mathcal{P}$. Finding the maximal reduction datum containing p_0 might require iteration because the definition of a reduction datum is self-referential, in that it stipulates conditions on $\mathfrak{n}$ which are to hold on the level set of the momenta defined by $\mathfrak{n}$ itself. Define

$$\mathfrak{n}^0 \equiv \mathfrak{g}^{\mathrm{sH}}, \qquad \mu^0 \equiv (\iota_{\mathfrak{n}^0}^* J)(p_0), \qquad \mathcal{P}^0 \equiv (\iota_{\mathfrak{n}^0}^* J)^{-1}(\mu^0)$$

and iterate

$$\mathfrak{n}^{i+1} \equiv \{\xi \in \mathfrak{n}^i : \xi \text{ is semi-Hamiltonian on } \mathcal{P}^i\},$$
$$\mu^{i+1} \equiv (\iota^*_{\mathfrak{n}^i} J)(p_0),$$
$$\mathcal{P}^{i+1} \equiv (\iota^*_{\mathfrak{n}^i} J)^{-1}(\mu^{i+1}).$$

The iteration stops, say at $\hat{\mathfrak{n}}, \hat{\mu}, \hat{\mathcal{P}}$, in finitely many steps because the dimension of $\mathfrak{n}^i$ increases with i but is finite. Setting

$$\hat{\mathcal{H}} \equiv \{h \in \mathcal{H} : h \text{ is semi-symplectic on } \hat{\mathcal{P}} \text{ and } h\hat{\mu} = \hat{\mu}\}$$

retrieves the unique maximal reduction datum $(\hat{\mathcal{H}}, \hat{\mathfrak{n}}, \hat{\mu})$ containing p_0. The iteration should be a pre-processing step to the reduction procedure, which will now be defined.

Let $R \equiv (\mathcal{H}, \mathfrak{n}, \mu)$ be a reduction datum, and define

$$\mathcal{P}_R \equiv (\iota^*_{\mathfrak{n}\cap\mathfrak{h}} J)^{-1}(\mu) = \{p \in \mathcal{P} : J_\xi(p) = \langle \mu, \xi\rangle \text{ for all } \xi \in \mathfrak{n}\cap\mathfrak{h}\}.$$

$\mathcal{H}$ acts on $\mathcal{P}_R$ by equivariance of J. Generally, for a subbundle $\mathcal{F}$ of $\mathcal{K}$, define

$$\mathcal{F}^{\omega\perp} = \{v \in \mathcal{K} : \omega(v, w) = 0 \text{ for all } w \in \mathcal{F}\};$$

in particular, $\mathcal{F}^{\omega\perp} \subseteq \mathcal{K}$ by definition, even if ω is defined on a subbundle strictly containing $\mathcal{K}$. Assume that

1. μ is a regular value of $\iota^*_{\mathfrak{n}\cap\mathfrak{h}} J$;
2. there is a smooth (onto submersive) quotient $\pi_R : \mathcal{P}_R \to \mathcal{P}_R/\mathcal{H}$;
3. $\mathcal{K}_R \equiv (\ker \boldsymbol{T}\pi_R \cap \mathcal{K})^{\omega\perp} \cap \boldsymbol{T}\mathcal{P}_R$ is a distribution;
4. $\ker \boldsymbol{T}\pi_R \cap \mathcal{K}_R$ is a distribution.

$(\ker \boldsymbol{T}\pi_R \cap \mathcal{K})^{\omega\perp} \subset \mathcal{K}$ by definition of ${}^{\omega\perp}$, so $\mathcal{K}_R \subseteq \mathcal{K}$. Define $\bar{\mathcal{K}}_R \equiv \boldsymbol{T}\pi_R\mathcal{K}_R$, which is a distribution on $\mathcal{P}_R$ since $\mathcal{K}_R$ is $\mathcal{H}$ invariant and $\ker \boldsymbol{T}\pi_R \cap \mathcal{K}_R$ is assumed to be a distribution.

The two-form ω descends to a 2-form $\bar{\omega}_R$ on $\bar{\mathcal{K}}_R$: if $\boldsymbol{T}\pi_R(v_i) = \bar{v}$, where $v_i \in (\mathcal{K}_R)_p$, $i = 1, 2$, then $v_2 - v_1 \in \ker \boldsymbol{T}\pi_R \cap \mathcal{K}$ so $\omega(v_2 - v_1, w) = 0$ for any $w \in (\mathcal{K}_R)_p$. This and $\mathcal{H}$ invariance of ω implies $\bar{\omega}$ well defined by $\bar{\omega}_R(\boldsymbol{T}\pi_R(v), \boldsymbol{T}\pi_R(w)) = \omega(v, w)$.

The two-form $\bar{\omega}_R$ is nondegenerate: if $\bar{v} \in (\bar{\mathcal{K}}_R)_{\bar{p}}$ and $\bar{\omega}_R(\bar{v}, \bar{w}) = 0$ for all $\bar{w} \in (\bar{\mathcal{K}}_R)_{\bar{p}}$, then $\boldsymbol{T}\pi_R(v) = \bar{v}$ for some $v \in (\mathcal{K}_R)_p$ and $\omega(v, w) = 0$ for all $w \in (\mathcal{K}_R)_p$, so $v \in (\mathcal{K}_R)_p^{\omega\perp} \cap (\mathcal{K}_R)_p$. It is required to show that $\bar{v} = 0$, i.e. $v \in \ker \boldsymbol{T}\pi_R$. Well,

$$\begin{aligned}(\ker \boldsymbol{T}\mathcal{P}_R \cap \mathcal{K})_p &= \ker \boldsymbol{d}(\iota^*_{\mathfrak{n}\cap\mathfrak{h}} J)(p) \cap \mathcal{K}_p \\ &= \{w_p \in \mathcal{K}_p : \boldsymbol{d}J_\xi(p) w_p = 0 \text{ for all } \xi \in \mathfrak{n}\cap\mathfrak{h}\} \\ &= \{w_p \in \mathcal{K}_p : \omega(\xi p, w_p) = 0 \text{ for all } \xi \in \mathfrak{n}\cap\mathfrak{h}\} \\ &= \big((\mathfrak{n}\cap\mathfrak{h})p\big)^{\omega\perp},\end{aligned}$$

so that

$$\begin{aligned}\mathcal{K}_R^{\omega\perp} &= \left((\ker \boldsymbol{T}\pi_R \cap \mathcal{K})^{\omega\perp} \cap \boldsymbol{T}\mathcal{P}_R\right)^{\omega\perp}\\ &= \left((\ker \boldsymbol{T}\pi_R \cap \mathcal{K})^{\omega\perp} \cap (\boldsymbol{T}\mathcal{P}_R \cap \mathcal{K})\right)^{\omega\perp}\\ &= (\ker \boldsymbol{T}\pi_R \cap \mathcal{K}) + (\boldsymbol{T}\mathcal{P}_R \cap \mathcal{K})^{\omega\perp}\\ &= (\ker \boldsymbol{T}\pi_R \cap \mathcal{K}) + (\mathfrak{n} \cap \mathfrak{h})\mathcal{P}_R\\ &\subseteq \mathfrak{h}\mathcal{P}_R\end{aligned}$$

and $v \in (\mathcal{K}_R)_p^{\omega\perp} \cap (\mathcal{K}_R)_p \subseteq \mathfrak{h}p \cap \boldsymbol{T}_p\mathcal{P}_R$. But $\bar{v} = \boldsymbol{T}\pi_R(v) = 0$ because $\mathfrak{h}p \cap \boldsymbol{T}_p\mathcal{P}_R \subseteq \ker \boldsymbol{T}_p\pi_R$, which is true as follows: if $\xi \in \mathfrak{h}$, $\xi p \in \boldsymbol{T}_p\mathcal{P}_R$, and $\eta \in \mathfrak{n} \cap \mathfrak{h}$, then

$$0 = \boldsymbol{d}J_\eta(\xi p) = \frac{d}{dt}\bigg|_{t=0} J_\eta\big(\exp(\xi t)p\big) = \frac{d}{dt}\bigg|_{t=0} \langle \exp(\xi t)\mu, \eta\rangle = \langle \operatorname{coad}_\xi \mu, \eta\rangle,$$

i.e. $\operatorname{coad}_\xi \mu$ annihilates $\mathfrak{n} \cap \mathfrak{h}$, so

$$\begin{aligned}\frac{d}{dt}\langle \exp(\xi t)\mu, \eta\rangle &= \frac{d}{ds}\bigg|_{s=0} \langle \exp(\xi t)\exp(\xi s)\mu, \eta\rangle\\ &= \frac{d}{ds}\bigg|_{s=0} \langle \exp(\xi s)\mu, \exp(-\xi t)\eta\rangle\\ &= \langle \operatorname{coad}_\xi \mu, \exp(-\xi t)\eta\rangle\\ &= 0,\end{aligned}$$

because $\exp(-\xi t)\eta \in \mathfrak{n} \cap \mathfrak{h}$. So $\exp(\xi t)\mu = \mu$, i.e. $\exp(\xi t) \in \mathcal{H}$ from which $\xi p \in \ker \boldsymbol{T}_p\pi_R$ as required.

The reduction has not yet accounted for conserved quantities that arise when $\mathfrak{n} \cap \mathfrak{h}$ is strictly included in $\mathfrak{n}$. Define

$$\mathcal{K}_R^\circ = \{\alpha | \boldsymbol{T}_p\mathcal{P}_R : p \in \mathcal{P}_R,\ \alpha \in \mathcal{K}_p^\circ,\ \langle \alpha, \mathfrak{g}p\rangle = 0\}$$

and assume this is a distribution. $\mathcal{K}_R^\circ$ is invariant under the action of $\mathcal{H}$ because $\mathcal{K}^\circ$ is ($\mathcal{H}$ acts semi-symplectically). Since $\mathcal{K}_R^\circ$ then annihilates $\ker \boldsymbol{T}\pi_R$, it descends to a distribution $\bar{\mathcal{K}}_R^\circ$ of $\mathcal{P}_R/\mathcal{H}$.

The *reduced space* is the semi-symplectic manifold $(\bar{\mathcal{P}}_R, \bar{\mathcal{K}}_R, \bar{\omega}_R, \bar{\mathcal{K}}_R^\circ)$ where $\bar{\mathcal{P}}_R \equiv \mathcal{P}_R/\mathcal{H}$. If $H : \mathcal{P} \to \mathbb{R}$ is $\mathcal{G}$ invariant then it descends to a function $\bar{H} : \bar{\mathcal{P}}_R \to \mathbb{R}$. The *reduced system* is defined to be semi-Hamiltonian system $(\bar{\mathcal{P}}_R, \bar{\mathcal{K}}_R, \bar{\omega}_R, \bar{\mathcal{K}}_R^0, \bar{H})$.

The semi-Hamiltonian vector field Y_H is tangent to $\mathcal{P}_R$ because, if $p \in \mathcal{P}_R$ and $\xi \in \mathfrak{n}$, then

$$\boldsymbol{d}J_\xi(p)Y_H(p) = \frac{dJ_\xi}{dt}(p) = 0$$

by the momentum equation. If $v_p \in \ker \boldsymbol{T}_p\pi_R \cap \mathcal{K}$, then

$$\omega\big(Y_H(p), v_p\big) = \boldsymbol{d}H(p)v_p = 0,$$

because H is $\mathcal{G}$ invariant. Thus $Y_H(p) \in (\ker \boldsymbol{T}_p\pi_R \cap K)^{\omega\perp}$, so $Y_H|\mathcal{P}_R$ takes values in $\mathcal{K}_R$. If $v_p \in \mathcal{K}_R$, then

$$\bar{\omega}_R\big(\boldsymbol{T}\pi_R Y_H(p), \boldsymbol{T}\pi_R v_p\big) = \omega\big(Y_H(p), v_p\big)$$
$$= \langle \boldsymbol{d}H, v_p\rangle = \langle \boldsymbol{d}(\bar{H}\pi_R), v_p\rangle = \langle \boldsymbol{d}\bar{H}, \boldsymbol{T}\pi_R v_p\rangle,$$

which shows that

$$\boldsymbol{T}\pi_R Y_H|\mathcal{P}_R = Y_{\bar{H}_R}\pi_R,$$

so the dynamics of the original semi-symplectic system, restricted to $\mathcal{P}_R$, intertwines the dynamics of the reduced semi-symplectic system. This completes the construction of the reduced system.

$\bar{\omega}_R$ is not defined except on $\bar{\mathcal{K}}_R$ because only on the distribution $\mathcal{K}_R$ does ω appropriately annihilate the vectors in $\ker \boldsymbol{T}\pi_R$, and without this ω will not drop to the quotient. The reduction cannot be expected to yield a semi-symplectic form defined anywhere except on the semi-symplectic distribution, even if the original semi-symplectic form ω is a two-form on the whole of $\mathcal{P}$. It would not be natural, in the definition of a semi-Hamiltonian systems, to enforce that ω be a restriction of a two-form on the whole phase space (as happens for nonholonomic systems defined by the Lagrange–d'Alembert principle) because then the semi-symplectic category would not close under reduction.

5.3. Example: the nonholonomic particle

The nonholonomic free particle is defined by the data

$$\begin{aligned}
Q &= \mathbb{R}^3 = \{q\} = \{(x, y, z)\},\\
\boldsymbol{T}Q &= \mathbb{R}^3 \times \mathbb{R}^3 = \{(q, v)\} = \{(x, y, z, v_x, v_y, v_z)\},\\
L &= \frac{1}{2}|v|^2 = \frac{1}{2}(v_x^2 + v_y^2 + v_z^2) - V(x, y, z),\\
\mathcal{D} &= \{(q, v) : v_z = a(y)v_x\}.
\end{aligned}$$

i.e. a free particle in three dimensions with the constraint $v_z = a(y)v_x$. The nonholonomic phase space is

$$\mathcal{P} = \boldsymbol{T}\mathcal{D} = \{(x, y, z, v_x, v_y)\},$$

so $\dim \mathcal{P} = 5$. This system appears (without any potential) in [4] who attribute it to [30], see also [2]. The system with the potential mentioned, and the orphaned momentum, are considered in [3]. The system is genuinely nonholonomic because $\mathcal{D}$ is spanned by

$$\frac{\partial}{\partial y}, \qquad \frac{\partial}{\partial x} + a(y)\frac{\partial}{\partial z},$$

and the Lie bracket of these is

$$\left(\frac{\partial}{\partial y}\right)\left(\frac{\partial}{\partial x} + a(y)\frac{\partial}{\partial z}\right) - \left(\frac{\partial}{\partial x} + a(y)\frac{\partial}{\partial z}\right)\left(\frac{\partial}{\partial y}\right) = a'(y)\frac{\partial}{\partial z},$$

which is nonzero everywhere that $a'(y) \neq 0$.

Semi-symplectic equations. The Lagrange–d'Alembert equations are the constraint, $dq/dt = v$ and $\delta L = 0$ on any vector in $\mathcal{D}$, where

$$\begin{aligned}\delta L &= \left(\frac{\partial L}{\partial q^i} - \frac{\partial^2 L}{\partial \dot{q}^i \partial q^j}\dot{q}^j - \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}\ddot{q}^j\right)\boldsymbol{d}q^i \\ &= -\left(\frac{\partial V}{\partial x} + \dot{v}_x\right)\boldsymbol{d}x - \left(\frac{\partial V}{\partial y} + \dot{v}_y\right)\boldsymbol{d}y - \left(\frac{\partial V}{\partial z} + \dot{v}_z\right)\boldsymbol{d}z,\end{aligned}$$

so

$$\begin{aligned}&\frac{dx}{dt} = v_x, \qquad \frac{dy}{dt} = v_y, \qquad \frac{dz}{dt} = v_z, \\ &\frac{dv_y}{dt} = -\frac{\partial V}{\partial y}, \qquad \frac{dv_x}{dt} + a(y)\frac{dv_z}{dt} = -\frac{\partial V}{\partial x} - a(y)\frac{\partial V}{\partial z}, \qquad v_z = a(y)v_x.\end{aligned}$$

This can be converted to a differential equation on $\mathcal{P}$ by eliminating v_z as follows:

$$\begin{aligned}\frac{dv_z}{dt} &= a'(y)v_x\frac{dy}{dt} + a(y)\frac{dv_x}{dt}, \\ \frac{dv_x}{dt} + a(y)\frac{dv_z}{dt} &= \left(1 + a(y)^2\right)\frac{dv_x}{dt} + a(y)a'(y)v_x v_y,\end{aligned}$$

giving the equations

$$\begin{aligned}&\frac{dx}{dt} = v_x, \qquad \frac{dy}{dt} = v_y, \qquad \frac{dz}{dt} = a(y)v_x, \\ &\frac{dv_x}{dt} = -\frac{1}{1 + a(y)^2}\left(a(y)a'(y)v_x v_y + \frac{\partial V}{\partial x} + a(y)\frac{\partial V}{\partial z}\right), \\ &\frac{dv_y}{dt} = -\frac{\partial V}{\partial y}.\end{aligned}$$

Since

$$\tau_Q|\mathcal{P}(x, y, z, v_x, v_y) = (x, y, z), \qquad \mathcal{K} = (\boldsymbol{T}\tau_Q|\mathcal{P})^{-1}\mathcal{D},$$

$\mathcal{K}$ is spanned by

$$K_1 = \frac{\partial}{\partial x} + a(y)\frac{\partial}{\partial z}, \qquad K_2 = \frac{\partial}{\partial y}, \qquad K_3 = \frac{\partial}{\partial v_x}, \qquad K_4 = \frac{\partial}{\partial v_y}.$$

Pulling ω_L back to $\mathcal{P}$ by the substitution $v_z = a(y)v_x$ gives

$$\begin{aligned}\omega_L &= \boldsymbol{d}x \wedge \boldsymbol{d}v_x + \boldsymbol{d}y \wedge \boldsymbol{d}v_y + \boldsymbol{d}z \wedge \boldsymbol{d}v_z \\ &= \boldsymbol{d}x \wedge \boldsymbol{d}v_x + \boldsymbol{d}y \wedge \boldsymbol{d}v_y + \boldsymbol{d}z \wedge (a'(y)v_x\boldsymbol{d}y + a(y)\boldsymbol{d}v_x),\end{aligned}$$

while the energy pulls back to

$$E = \frac{1}{2}\left(1 + a(y)^2\right)v_x^2 + \frac{1}{2}v_y^2 + V(x, y, z).$$

With respect to the basis K_i, i.e. written in terms of the dual basis K^i,

$$\omega_L|\mathcal{K}\times\mathcal{K} = a(y)a'(y)v_x K^1\wedge K^2 + \big(1+a(y)^2\big)K^1\wedge K^3 + K^2\wedge K^4.$$

If $f(x,y,z,v_x,v_y)$ is any smooth function then

$$\boldsymbol{d}f|\mathcal{K} = \left(\frac{\partial f}{\partial x} + a(y)\frac{\partial f}{\partial z}\right)K^1 + \frac{\partial f}{\partial y}K_2 + \frac{\partial f}{\partial v_x}K_3 + \frac{\partial f}{\partial v_y}K_4,$$

and if $Y_f = Y^i K_i$ then

$$\begin{aligned}\boldsymbol{i}_{Y_f}\omega_L &= \big(-a(y)a'(y)v_x Y^2 - \big(1+a(y)^2\big)Y^3\big)K^1\\ &\quad + \big(a(y)a'(y)v_x Y^1 - Y^4\big)K^2 + \big(1+a(y)^2\big)Y^1K^3 + Y^2K^4,\end{aligned}$$

from which it is easily computed that

$$\begin{aligned}Y_f &= \frac{1}{1+a(y)^2}\frac{\partial f}{\partial v_x}K_1 + \frac{\partial f}{\partial v_y}K_2\\ &\quad - \frac{1}{1+a(y)^2}\left(\frac{\partial f}{\partial x} + a(y)\frac{\partial f}{\partial z} + a(y)a'(y)v_x\frac{\partial f}{\partial v_y}\right)K_3\\ &\quad + \left(\frac{a(y)a'(y)v_x}{1+a(y)^2}\frac{\partial f}{\partial v_x} - \frac{\partial f}{\partial y}\right)K_4.\end{aligned}$$

Substituting E for f gives

$$\begin{aligned}Y_E &= v_xK_1 + v_yK_2\\ &\quad - \frac{1}{1+a(y)^2}\left(a(y)a'(y)v_xv_y + \frac{\partial V}{\partial x} + a(y)\frac{\partial V}{\partial z}\right)K_3 - \frac{\partial V}{\partial y}K_4\end{aligned}$$

which after substitution of the K_i gives the same equations as derived directly from the variational principle. This verifies the semi-symplectic equations.

Curvature, holonomic subsystems. Any complement to $\mathcal{K}$ can be used to obtain ν and compute the curvature, and the numerical value of the curvature depends on the complement and its basis. An obvious complement is $\partial/\partial z$ giving

$$\nu = -a(y)\boldsymbol{d}x + \boldsymbol{d}z, \qquad \Delta_{\mathcal{K}} = -\boldsymbol{d}\nu = a'(y)\boldsymbol{d}x\wedge\boldsymbol{d}y.$$

Suppose y_0 is such that $a'(y_0)=0$ and $\partial V/\partial y(x,y_0,z)=0$ for all x,z. Let

$$\mathcal{P}_{y_0} = \{(x,y,z,v_x,v_y) : y=y_0, v_y=0\}.$$

Putting $a(y_0)\equiv a_0$ into the equations of motion gives

$$\frac{dx}{dt} = v_x, \qquad \frac{dz}{dt} = a_0v_x, \qquad \frac{dv_x}{dt} = -\frac{1}{1+a_0^2}\left(\frac{\partial V}{\partial x} + a_0\frac{\partial V}{\partial z}\right).$$

Solve for z as follows:

$$\frac{dz}{dt} - a_0 v_x = \frac{dz}{dt} - a_0 \frac{dx}{dt} = \frac{d}{dt}(z - a_0 x) = 0,$$

so $z = a_0 x + C$ where C is a constant. So the evolution is equivalent to this equation and the Hamiltonian evolution

$$\frac{dx}{dt} = v_x, \qquad \frac{dv_x}{dt} = -\frac{\partial \tilde{V}}{\partial x},$$

where $\tilde{V}(x) = V(x, y_0, a_0 x + C)/(1 + a_0^2)$. Let the flow of x, v_x system be

$$F_t(x, v_x) = \big(x_t(x, v_x), v_{x,t}(x, v_x)\big),$$

so that the flow on $\mathcal{P}_{y_0}$ is

$$\tilde{x} = x_t(x, v_x), \qquad \tilde{y} = y, \qquad \tilde{z} = a_0(\tilde{x} - x) + z, \qquad \tilde{v}_x = v_{x,t}(x, v_x), \qquad \tilde{v}_y = 0.$$

$\mathcal{K}_{y_0} \equiv \boldsymbol{T}\mathcal{P}_{y_0} \cap \mathcal{K}$ is flow invariant because it is the annihilator on $T\mathcal{P}_{y_0}$ of $a_0 \boldsymbol{d}x - \boldsymbol{d}z$ and

$$a_0 \boldsymbol{d}\tilde{x} - \boldsymbol{d}\tilde{z} = a_0 \boldsymbol{d}\tilde{x} - \boldsymbol{d}\big(a_0(\tilde{x} - x) + z\big) = a_0 \boldsymbol{d}x - d\tilde{z},$$

and $\Delta_{T\mathcal{P}_{y_0}} = 0$ on $\boldsymbol{T}\mathcal{P}_{y_0}$, and ω_L is closed, so Theorem 5.2 implies that flow is symplectic on $\boldsymbol{T}\mathcal{P}_{y_0} \cap \mathcal{K}$.

ω_L pulled back to $\mathcal{P}_{y_0}$ is

$$\omega_L = \boldsymbol{d}x \wedge \boldsymbol{d}v_x + a_0 \boldsymbol{d}z \wedge \boldsymbol{d}v_x$$

and using symplecticity of the (Hamiltonian) flow in x, v_x, the pull-back of ω_L by the flow is

$$\begin{aligned}
\tilde{\omega}_L &= \boldsymbol{d}\tilde{x} \wedge \boldsymbol{d}\tilde{v}_x + a_0 \boldsymbol{d}\tilde{z} \wedge \boldsymbol{d}\tilde{v}_x \\
&= \boldsymbol{d}x \wedge \boldsymbol{d}v_x + a_0(a_0 \boldsymbol{d}\tilde{x} - a_0 \boldsymbol{d}x + \boldsymbol{d}z) \wedge \boldsymbol{d}\tilde{v}_x \\
&= \boldsymbol{d}x \wedge \boldsymbol{d}v_x + a_0^2 \boldsymbol{d}x \wedge \boldsymbol{d}v_x + a_0(\boldsymbol{d}z - a_0 \boldsymbol{d}x) \wedge \boldsymbol{d}\tilde{v}_x.
\end{aligned}$$

The difference between the pull-back of ω_L by the flow and ω_L itself is

$$\begin{aligned}
\tilde{\omega}_L - \omega_L &= a_0^2 \boldsymbol{d}x \wedge \boldsymbol{d}v_x + a_0(\boldsymbol{d}z - a_0 \boldsymbol{d}x) \wedge \boldsymbol{d}\tilde{v}_x - a_0 \boldsymbol{d}z \wedge \boldsymbol{d}v_x \\
&= a_0(\boldsymbol{d}z - a_0 \boldsymbol{d}x) \wedge (\boldsymbol{d}\tilde{v}_x - \boldsymbol{d}v_x)
\end{aligned}$$

and the factorization shows that $\tilde{\omega}_L = \omega_L$ on $\mathcal{K}$ because $\boldsymbol{d}z - a_0 \boldsymbol{d}x$ annihilates $\mathcal{K}_{y_0}$. This verifies that the flow is symplectic on $\mathcal{K}_{y_0}$.

Symmetry, reduction. The nonholonomic phase space $\mathcal{P}$ is not invariant under the action of $\mathbb{R}^3$ by translations in x, y, z because

$$(t_1, t_2, t_3)\big(x, y, z, v_x, v_y, a(y)v_x\big) = \big(x + t_1, y + t_2, z + t_3, v_x, v_y, a(y)v_x\big)$$

but the latter should be $a(y + t_2)v_x$. So only the sub-action of translations in x, y is present in the semi-symplectic formalism. Assume the largest available symmetry:

$V = V(y)$, i.e. V is a function of y, and $\mathcal{G} = \mathbb{R}^2$ acting by translations in x and y on $\mathcal{Q}$. This has momentum map $J = (v_x, v_z)$ but the momentum is not conserved because the generators are not semi-Hamiltonian. Invariant functions on $\mathcal{P}$ are of the form $f(y, v_x, v_y)$, give differentials spanned by K^2, K^3, K^4, so $\mathcal{K}^\circ$ is the annihilator of K_1, and give semi-Hamiltonian vector fields

$$Y_f = \frac{1}{1+a(y)^2}\frac{\partial f}{\partial v_x}K_1 + \frac{\partial f}{\partial v_y}K_2 - \frac{a(y)a'(y)v_x}{1+a(y)^2}\frac{\partial f}{\partial v_y}K_3 + \left(\frac{a(y)a'(y)v_x}{1+a(y)^2}\frac{\partial f}{\partial v_x} - \frac{\partial f}{\partial y}\right)K_4,$$

so

$$\mathcal{K}^\bullet = \operatorname{span}\big\{K^1, K^4, a(y)a'(y)v_xK^3 - \big(1+a(y)^2\big)K^2\big\}.$$

The differential of the function $j \equiv v_x^2\big(1+a(y)^2\big)$ is annihilated by this so $\mathcal{K}^\bullet$ is integrable and there is this one orphaned conserved quantity.

The momentum v_y is not preserved as is evident from the presence of $V(y)$ in the dv_y/dy evolution equation. If the Lagrangian is $\mathbb{R}^3$ invariant, i.e. $V(y) = 0$ then the momentum v_y is conserved by direct observation of the evolution equations, but this is not predicted by the general theory because translation in y still does not preserve the nonholonomic phase space. Understanding this requires the analysis of the orphaned conserved quantity: $v_y^2/2 + V(y)$ is conserved because it happens to be the energy minus half the orphaned $j = v_x^2\big(1+a(y)^2\big)$. If $V(y) = 0$ then this implies v_y is conserved.

Pick any $p_0 \equiv (0, y_0, 0, v_{x,0}, v_{y,0})$; the objective is to construct the reduced space through p_0. Since none of the generators are semi-Hamiltonian, one uses the reduction datum $R = (\mathcal{G}, 0, 0)$, i.e. $\mu_0 = 0$ as an element of $\mathfrak{n}^*$ where $\mathfrak{n} = 0$. There is no pre-processing step since there are no conserved momenta.

There is no momentum level set so the reduced space is

$$\bar{\mathcal{P}}_R = \mathcal{P}_R/\mathbb{R}^2 = \{(x, y, z, v_x, v_y) : x = z = 0\} = \big\{(y, v_x, v_y)\big\}$$

irrespective of the choice of p_0. The distribution $\mathcal{K}_R$ is

$$\begin{aligned}\mathcal{K}_R &= \big(\ker \boldsymbol{T}\pi_R \cap \mathcal{K}\big)^{\omega_L\perp} \cap \boldsymbol{T}\mathcal{P}_R\\ &= (\mathbb{R}K_1)^{\omega_L\perp}\\ &= \operatorname{ann}(\boldsymbol{i}_{K_1}\omega_L)\\ &= \operatorname{ann}\big(a(y)a'(y)v_xK^2 + \big(1+a(y)^2\big)K^3\big)\\ &= \operatorname{span}\big\{K^1, K^4, a(y)a'(y)v_xK^3 - \big(1+a(y)^2\big)K^2\big\}\end{aligned}$$

and $\boldsymbol{T}\pi_R$ of this is the semi-symplectic distribution of the reduced space,

$$\bar{\mathcal{K}}_R = \operatorname{span}\big\{\boldsymbol{T}\pi_RK^4, a(y)a'(y)v_x\boldsymbol{T}\pi_RK^3 - \big(1+a(y)^2\big)\boldsymbol{T}\pi_RK^2\big\},$$

i.e. the two-dimensional distribution of $\bar{\mathcal{P}}_R = \{(y, v_x, v_y)\}$ spanned by

$$\bar{K}_1 \equiv a(y)a'(y)v_x \frac{\partial}{\partial v_x} - (1 + a(y)^2)\frac{\partial}{\partial y}, \qquad \bar{K}_2 \equiv \frac{\partial}{\partial v_y}.$$

With dual basis $\bar{K}^1, \bar{K}^2$ this gives

$$\bar{\omega}_R = -(1 + a(y)^2)\bar{K}^1 \wedge \bar{K}^2, \qquad \bar{H}_R = \frac{1}{2}(1 + a(y)^2)v_x^2 + \frac{1}{2}v_y^2 + V(y).$$

Amusingly, $\bar{\mathcal{K}}_R$ is integrable and it annihilates the push-down $\bar{j} \equiv v_x^2(1 + a(y)^2)$. The distribution $\bar{\mathcal{K}}^\circ$, found by evaluating the one-forms that annihilate K_1 on vectors in $\mathcal{P}_R$ which project to given vectors in $\bar{\mathcal{P}}_R$, is just the whole of $T\bar{\mathcal{P}}_R$. The corresponding $\bar{\mathcal{K}}^\bullet$ is the same as the semi-symplectic distribution $\bar{\mathcal{K}}_R$.

So the reduction of this nonholonomic system is symplectic. The symplectic systems can be found by eliminating v_x using

$$v_x^2 = \frac{C}{1 + a(y)^2},$$

i.e. pull back the form $\bar{\omega}_R$ and the Hamiltonian $\bar{H}_R$ by the map

$$(y, v_y) \mapsto (y, v_y, v_x), \qquad v_x^2 = \frac{C}{1 + a(y)^2}.$$

By implicit differentiation, this sends $\partial/\partial y$ to

$$\frac{\partial}{\partial y} - \frac{Ca(y)a'(y)}{v_x(1 + a(y)^2)^2}\frac{\partial}{\partial v_x} = \frac{\partial}{\partial y} - \frac{a(y)a'(y)v_x}{(1 + a(y)^2)}\frac{\partial}{\partial v_x} = -\frac{1}{1 + a(y)^2}\bar{K}_1$$

and $\partial/\partial v_y$ to $\bar{K}_2$, so the pull back of $\bar{\omega}_R$ is the canonical form $\boldsymbol{d}y \wedge \boldsymbol{d}v_y$ and the pull back of the Hamiltonian is $v_y^2/2 + V(y) + C/2$. This is a single particle of unit mass moving in the potential $V(y)$. The reduced space consists of a trivial one-parameter family of single particles in a potential $V(y)$, and the reduction simply isolates the obvious Hamiltonian evolution in the y variables.

6. Nonlinear constraints

Let $\mathcal{C} \subseteq \boldsymbol{T}\mathcal{Q}$ be a submanifold such that $\tau_{\mathcal{Q}}$ is a submersion on $\mathcal{C}$, and let $\mathcal{W}$ be a subbundle of the pull-back bundle $(\tau_{\mathcal{Q}}|\mathcal{C})^*(\boldsymbol{T}\mathcal{Q})$. As stated in Introduction, the Lagrange–d'Alembert principle for this data is

$$\begin{aligned} &\boldsymbol{d}S(q(t)) \cdot \delta q(t) = 0 \qquad \text{for all} \quad \delta q(t) \in \mathcal{W}_{q'(t)}, \\ &q'(t) \in \mathcal{C}. \end{aligned} \tag{18}$$

Since $\tau_{\mathcal{Q}}$ is assumed to be a submersion on $\mathcal{C}$, one choice is *Chetaev's rule* $\mathcal{W} = \dot{\mathcal{C}}$ where

$$\dot{\mathcal{C}}_{vq} \equiv \{w_q \in \mathcal{C} : \mathrm{vert}_{v_q} w_q \in \boldsymbol{T}\mathcal{C}\},$$

but other choices might be appropriate, as discussed in [23]. If $\mathcal{C}$ is a distribution on $\boldsymbol{T}\mathcal{Q}$ then $\dot{\mathcal{C}}_{v_q} = \mathcal{C}_q$ and Chetaev's rule gives a linearly constrained system as already considered. Writing the first of (18) in Lagrange multiplier form gives $\delta L = -\lambda$ where λ is in the annihilator of $\mathcal{W}$. Thus $\mathcal{W}$ has the physical interpretation that the annihilator of $\mathcal{W}_{v_q}$ is the vector space of the constraint forces at state v_q.

DEFINITION 6.1. A Lagrangian L is *$(\mathcal{C}, \mathcal{W})$-regular* if, for all $v_q \in \mathcal{C}$, the bilinear map $\boldsymbol{F}^2L(v_q)|\dot{\mathcal{C}}_{v_q} \times \mathcal{W}_{v_q}$ is nonsingular, i.e. if $u_q \in \dot{\mathcal{C}}_{v_q}$ and $\boldsymbol{F}^2L(v_q)(u_q, w_q) = 0$ for all $w_q \in \mathcal{W}_{v_q}$ then $u_q = 0$.

THEOREM 6.1. *If L is a C^k $(\mathcal{C}, \mathcal{W})$-regular Lagrangian, $k \geq 2$, and $\mathcal{C}$ is C^{k-2}, then there is a unique C^{k-2} second order vector field $Y_{\delta L}$ on $\mathcal{C}$ such that $\delta L \circ Y_{\delta L}(v_q) \in \operatorname{ann} \mathcal{W}_{v_q}$ for all $v_q \in \mathcal{C}$.*

Proof: The proof follows that of Theorem 3.2: by Theorem 3.1, Eqs. (18) are equivalent to

$$\begin{aligned} &\delta L\big(q''(t)\big)\delta q(t) = 0 \qquad \text{for all } \delta q \text{ such that } \delta q(t) \in \mathcal{W}_{q'(t)} \\ &q'(t) \in \mathcal{C}. \end{aligned} \tag{19}$$

In coordinates, $Y_{\delta L}$ has the form

$$Y_{\delta L}(q, \dot{q}) = (q, \dot{q}, \dot{q}, \ddot{q}),$$

where $\ddot{q}$ is determined from (19). Here $Y_{\delta L}(q, \dot{q}) \in \boldsymbol{T}\mathcal{C}$ so $(q, \ddot{q}) \in \dot{\mathcal{C}}_{(q,\dot{q})}$. Choosing any $(q, \dot{q}, \dot{q}, \ddot{q}_0) \in T\mathcal{C}$, the problem is reduced to determining

$$(q, \dot{q}, \dot{q}, \ddot{q}) - (q, \dot{q}, \dot{q}, \ddot{q}_0) = (q, \dot{q}, 0, \ddot{q} - \ddot{q}_0),$$

which is vertical. From this it is easily seen that (19) is a linear equation for $\ddot{q} - \ddot{q}_0$ with matrix nonsingular if $\boldsymbol{F}^2L(v_q)|\dot{\mathcal{C}}_{v_q} \times \mathcal{W}_{v_q}$ is nonsingular. □

Since the fiber dimension of $\dot{\mathcal{C}}$ is $\dim \mathcal{C} - \dim \mathcal{Q}$, the fiber dimension of $\mathcal{W}$ must also be this if L is $(\mathcal{C}, \mathcal{W})$-regular, in which case the codimension of the fibers of $\mathcal{W}$ is the same as the dimension of the constraint. Theorem 6.1 confirms the expectation that there ought to be the same dimension of constraint forces as the dimension of constraints under generic regularity conditions that also give proper equations of motion.

With the time evolution established, the action can be pulled back to the solution space to obtain Eqs. (11) and (12) i.e.

$$\boldsymbol{d}S\big(q(t)\big) \cdot w_{v_q} = (F_t^*\theta_L - \theta_L)w_{v_q} + \int_0^t \delta L\big(q''(t)\big) \cdot \delta q \, dt, \tag{20}$$

$$\delta q = \boldsymbol{T}\tau_{\mathcal{Q}} \, \boldsymbol{T}F_t w_{v_q} = \boldsymbol{T}(\tau_{\mathcal{Q}} \circ F_t)w_{v_q},$$

where F_t is the flow of $Y_{\delta L}$.

Suppose Lie group G acts on $\mathcal{Q}$ and that the lift of this action to $\boldsymbol{T}\mathcal{Q}$ leaves L and $\mathcal{C}$ invariant. If $\xi \in \mathfrak{g}$ then conservation of the momentum J_ξ along an evolution

$q(t)$ follows from substituting $w_{v_q} = \xi_{TQ}(v_q)$ into (20), if $\delta q(t) = \xi q(t) \in \mathcal{W}_{q'(t)}$ for all t. In absence on any other information, this can be guaranteed only by assuming $\xi q = \boldsymbol{T}\tau_Q \xi_{TQ} v_q \in \mathcal{W}_{v_q}$ for all $v_q \in \mathcal{C}_q$. This is equivalent to $q\xi \in \mathcal{D}$ where

$$\mathcal{D} = \bigcup_{q \in Q} \bigcap_{v_q \in \mathcal{C}_q} \mathcal{W}_{v_q},$$

i.e. $\mathcal{D}$ is the fiberwise union over Q of the intersections of the $\mathcal{W}_{v_q}$ as v_q ranges over $\mathcal{C}_q$. Whereas every "fiber" of $\mathcal{D}$ is a vector space, there is no reason to imagine that $\mathcal{D}$ is a vector bundle or even that it consists of anything but the zero section. Thus conservation of momentum for nonlinearly constrained systems occurs under rather more strict conditions than for linearly constrained systems. The nonholonomic momentum equation, however, along a solution $v(t) \equiv F_t(v_q)$, is valid under the assumption that $\xi_{TQ}\big(v(t)\big) \in \mathcal{W}_{v(t)}$, with an identical proof as in the case of linear constraints. Symplecticity in the case of nonlinear constraints is similar, in that Proposition 4.1 of Section 4.2 is true under the alteration that the distribution $\mathcal{D}$ is replaced by the subset $\mathcal{D}$ above.

To find the implications of the variational development for the equations of motion, differentiate (20) in t, as in Section 4.3, to obtain

$$\boldsymbol{i}_{Y_{\delta L}} \omega_L(w_{v_q}) = \boldsymbol{d}E(w_{v_q}) + \delta L \circ Y_{\delta L}\big(\boldsymbol{T}\tau_Q w_{v_q}\big).$$

The development structurally diverges from the case of linear constraints, because the last term vanishes only for $\boldsymbol{T}\tau_Q w_{v_q} \in \mathcal{W}_{v_q}$, whereas the principle condition on $Y_{\delta L}$ is not this but rather $\boldsymbol{T}\tau_Q Y_{\delta L} \in \mathcal{C}$. Define

$$\mathcal{K}_{\mathcal{C},\mathcal{W}} = \boldsymbol{T}\mathcal{C} \cap (\boldsymbol{T}\tau_Q)^{-1}\mathcal{W},$$

$$\mathcal{M}_{\mathcal{C}} = \boldsymbol{T}\mathcal{C} \cap (\boldsymbol{T}\tau_Q)^{-1}\mathcal{C}.$$

$\mathcal{K}_{\mathcal{C},W} \subseteq \boldsymbol{T}\mathcal{C}$ is a distribution and $\mathcal{M}_{\mathcal{C}} \subseteq \boldsymbol{T}\mathcal{C}$ is a submanifold, and

$$\boldsymbol{i}_{Y_{\delta L}} \omega_L(w_{v_q}) = \boldsymbol{d}E(w_{v_q}), \qquad w_{v_q} \in \mathcal{K}_{\mathcal{C},W}, \tag{21}$$

since $\delta L \circ Y_{\delta L}(v_q)$ takes values in $\operatorname{ann}\mathcal{W}_{v_q}$. *But $Y_{\delta L}$ is in $\mathcal{M}_{\mathcal{C}}$ rather than $\mathcal{K}_{\mathcal{C},W}$, so* (21) *is not semi-Hamiltonian.*

Irrespective of the regularity of the Lagrangian L, (21) does not determine $Y_{\delta L}$ in the directions of the ω_L orthogonal complement of $\mathcal{K}_{\mathcal{C},\mathcal{W}}$, i.e.

$$\mathcal{Z}_{\mathcal{C},\mathcal{W}} \equiv \{w_v \in \boldsymbol{T}\mathcal{C} : \omega_L(w_v, \tilde{w}_v) = 0 \text{ for all } \tilde{w}_v \in (\mathcal{K}_{\mathcal{C},W})_v\}.$$

Set $Z \equiv Y_{\delta L} - Y_E$ where Y_E is defined from the semi-symplectic equations as before, but using $\mathcal{K}_{\mathcal{C},W}$. Then

$$\boldsymbol{i}_Z \omega_L(w) = \boldsymbol{i}_{Y_{\delta L}} \omega_L(w) - \boldsymbol{i}_{Y_E} \omega_L(w) = \boldsymbol{d}E(w) - \boldsymbol{d}E(w) = 0$$

for all $w \in \mathcal{K}_{\mathcal{C},W}$, so that $Z \in \boldsymbol{T}\mathcal{C}$ and $Z \in \mathcal{Z}_{\mathcal{C},W}$. This can be abstracted as follows.

DEFINITION 6.2. A *null-forcing* of semi-Hamiltonian system $(\mathcal{P}, \omega, \mathcal{K}, H)$ is a vector field Z on $\mathcal{P}$ such that $\boldsymbol{i}_Z \omega \in \operatorname{ann}\mathcal{K}$. $p(t) \in \mathcal{P}$ is a *time evolution* of the null-forced system if it is an integral curve of the vector field $Y_H + Z$.

The definition is non-vacuous only when the semi-symplectic form is defined on a distribution strictly containing $\mathcal{K}$; otherwise nondegeneracy of ω implies $Z = 0$ necessarily. What has been demonstrated above is that a nonholonomic system $L : TQ \to \mathbb{R}$ with (nonlinear) constraint submanifold $\mathcal{C} \subseteq TQ$ such that 1) τ_Q is a submersion on $\mathcal{C}$, and 2) L is $(\mathcal{C}, \mathcal{W})$-regular, is naturally a null-forced semi-symplectic system, such that the semi-Hamiltonian form is defined on all of the tangent space of phase space.

The null-forcing Z is provided by the variational principle and does not seem to belong to the semi-symplectic category. An important structural issue, unresolved at this time, is whether or not Z inherits some properties from the variational principle or is essentially arbitrary. In other words, when investigating null-forced semi-Hamiltonian systems in general, is it physically legitimate to posit an unrestricted null-forcing? The problem is to, in some appropriate sense, answer whether there is a Lagrangian L corresponding to an apriori given Z.

That the equations of motion cannot be expressed in terms of the almost-Poisson bracket in the case of nonlinear constraints is shown by [8]. Leibniz brackets are used in [28] to write the equations of motion in the case of affine constraints. For a null-forced Hamiltonian system it would be natural to use the bracket

$$[f, g] = \{f, g\} + Z(f),$$

where $\{,\}$ is the almost-Poisson bracket, but this is not a Leibniz bracket because it is not a derivation in its second slot. On the other hand, if $g \mapsto Z_g$ is a derivation, i.e. $Z_{g_1 g_2} = g_1 Z_{g_2} + g_2 Z_{g_1}$, and Z_g satisfies $Z_H = Z$, then the Leibniz bracket

$$[f, g] = \{f, g\} + Z_g(f)$$

would generate the motion.

From the perspective of the semi-Hamiltonian equations, conservation of energy, the proof of which follows directly from the antisymmetry of the semi-symplectic form, fails for nonlinearly constrained systems because the disparity between $\mathcal{M}_\mathcal{C}$, which is bound to the first slot of ω_L through $Y_{\delta L}$, and $\mathcal{K}_{\mathcal{C},W}$, which is bound to the second slot. Since $Y_{\delta L}$ is second order, $T\tau_Q Y_{\delta L}(v_q) = v_q \in \mathcal{K}_{\mathcal{C},\mathcal{W}}$ if and only if $v_q \in \mathcal{W}_{v_q}$. Thus $Z(v_q) = 0$ if and only if $v_q \in \mathcal{W}_{v_q}$, so the system is semi-Hamiltonian if $\mathcal{W}$ contains the diagonal. This is understandable since (10) implies

$$\frac{dE}{dt} = -\boldsymbol{i}_{Y_{\delta L}}(\boldsymbol{i}_{Y_{\delta L}}\omega_L - \boldsymbol{d}E) = -\delta L \circ Y_{\delta L} T\tau_Q Y_{\delta L} = \lambda(v_q)(v_q),$$

where λ is the constraint force. So, as expected, the rate of change of energy is power done by the constraint force (i.e. its contraction with velocity). If the constraint forces at v_q necessarily annihilate v_q, then they can do no work and the system conserves energy and is semi-symplectic. As is easily verified, if $\mathcal{W}$ is obtained by Chetaev's rule then it contains the diagonal if and only if the Liouville vector field $v_q \mapsto \mathrm{vert}_{v_q} v_q$ is tangent to $\mathcal{C}$, so these results generalize Propositions 3.3 and 3.8 of [9] to the situation where $\mathcal{W}$ is not necessarily obtained from Chetaev's rule.

Theorems 5.1 and 5.2 remain true with the semi-Hamiltonian evolution vector field replaced by the null-forced evolution vector field, because within the semi-symplectic form, the null-forcing Z annihilates any vector field with values in $\mathcal{K}$. Thus null-forced semi-symplectic systems have remnants of the symplectic category under the same conditions as semi-symplectic systems do.

Suppose a Lie group $\mathcal{G}$ acts semi-symplectically on $(\mathcal{P}, \omega, \mathcal{K})$, that H is $\mathcal{G}$ invariant, that Z is equivariant, and that J is a momentum mapping. By the conventions of Definition 5.6, $J_\xi(v_p) = (\boldsymbol{i}_{\xi_\mathcal{P}}\omega)(v_p)$ for all v_p in the distribution on which ω is defined. In particular, for $\xi \in \mathfrak{g}$ and $p \in \mathcal{P}$ such that $\xi p \in \mathcal{K}$,

$$\langle J_\xi(p), Z(p)\rangle = \omega\big((\xi p, Z(p)\big) = 0,$$

so that the momentum equation (Theorem 5.3) is true without change for null-forced semi-symplectic systems.

REFERENCES

[1] R. Abraham and J. E. Marsden: *Foundations of Mechanics*, Addision-Wesley, second edition, 1978.

[2] L. Bates: Examples of singular nonholonomic reduction, *Rep. Math. Phys.* **42** (1998), 231–247.

[3] L. Bates, H. Graumann and G. MacDonnell: Examples of gauge conservation laws in nonholonomic mechanics, *Rep. Math. Phys.* **37** (1996), 295–308.

[4] L. Bates and J. Śniatycki: Nonholonomic reduction, *Rep. Math. Phys.* **32** (1993), 99–115.

[5] G. Blankenstein and A. J. van der Schaft: Symmetry and reduction in implicit generalized Hamiltonian systems, *Rep. Math. Phys.* **47** (2001), 57–100.

[6] A. M. Bloch, P. S. Krishnaprasad, J. E. Marsden and R. M. Murray: Nonholonomic systems with symmetry, *Arch. Rational Mech. Anal.* **136** (1996), 21–99.

[7] A. M. Bloch, J. E. Marsden and D. V. Zenkov: Nonholonomic dynamics, *Notices Amer. Math. Soc.* **52** (2005), 324–333.

[8] F. Cantrijn, M. de León and D. M. de Diego: On almost-Poisson structures in nonholonomic mechanics, *Nonlinearity* **12** (1999), 721–737.

[9] F. Cantrijn, M. de León, J. C. Marrero and D. M. de Diego: Reduction of nonholonomic mechanical systems with symmetries, *Rep. Math. Phys.* **42** (1998), 25–45.

[10] F. Cantrijn, M. de León, J. C. Marrero and D. M. de Diego: Reduction of constrained systems with symmetries, *J. Math. Phys.* **40** (1999), 795–820.

[11] F. Cariñena and M. F. Rañada: Lagrangian systems with constraints: A geometric approach to the method of Lagrange multipliers, *J. Phys. A* **26** (1993), 1335–1351.

[12] H. Cendra, J. E. Marsden and T. S. Ratiu: Geometric mechanics, Lagrangian reduction, and nonholonomic systems, In B Enguist and W. Schmidt, editors, *Mathematics unlimited, 2001 and beyond*, Springer-Verlag 2001, pp. 221–273.

[13] J. Cortes and M. de León: Reduction and reconstruction of the dynamics nonholonomic systems, *J. Phys. A* **32** (1999), 8615–8645.

[14] R. Cushman, D. Kemppainen, J. Śniatycki and L. Bates: Geometry of nonholonomic constraints, *Rep. Math. Phys.* **36** (1995), 275–286.

[15] M. de León and D. M. de Diego: On the geometry non-holonomic Lagrangian systems, *J. Phys. A* **37** (1996), 3389–3414.

[16] M. de León, J. C. Marrero and D. M. de Diego: Mechanical systems with nonlinear constraints, *Internat. J. Theoret. Phys.* **36** (1997), 979–995.

[17] M. de León, J. C. Marrero and D. M. de Diego: Vakonomic mechanics versus non-holonomic mechanics: a unified geometrical approach, *J. Geom. Phys.* **35** (2000), 126–144.

[18] W. S. Koon and J. E. Marsden: The Hamiltonian and Lagrangian approaches to the dynamics of nonholonomic systems, *Rep. Math. Phys.* **40** (1997), 21–62.

[19] W. S. Koon and J. E. Marsden: Poisson reduction for nonholonomic mechanical systems with symmetry, *Rep. Math. Phys.* **42** (1998), 101–134.
[20] A. D. Lewis: Lifting distributions to tangent and jet bundles, unpublished, 1998.
[21] A. D. Lewis and R. M. Murray: Variational principles for constrained systems: theory and experiment, *Internat. J. Non-Linear Mechanics* **30** (1995), 793–815.
[22] C.-M. Marle: Reduction of constrained mechanical systems and stability of relative equilibria, *Commun. Math. Phys.* **174** (1995), 295–318.
[23] C. M. Marle: Various approaches to conservative and nonconservative nonholonomic systems, *Rep. Math. Phys.* **42** (1998), 211–229.
[24] G. Marmo, E.J. Saletan, A. Simoni and B. Vitale: *Dynamical systems*, Wiley, Chichester, 1985.
[25] J. E. Marsden, R. Montgomery and T. Ratiu: Reduction, symmetry, and phases in mechanics, *Mem. Amer. Math. Soc.* **88** (1990), 436.
[26] J. E. Marsden, G. W. Patrick and S. Shkoller: Multisymplectic geometry, variational integrators, and nonlinear PDEs, *Commun. Math. Phys.* **199** (1998), 351–395.
[27] J. E. Marsden and T. S. Ratiu: *Introduction to Mechanics and Symmetry*, vol. 17 of *Texts in Applied Mathematics*, Springer, 1994.
[28] J-P. Ortega and V. Planas-Biesa: Dynamics on Leibniz manifolds, *J. Geom. Phys.* **52** (2004), 1–27.
[29] J-P. Ortega and T. S. Ratiu: The optimal momentum map, In *Geometry, Mechanics, and Dynamics*, Springer, New York 2002, pp. 329–362.
[30] R. Rosenberg: *Analytical Dynamics*, Plenum Press, 1977.
[31] W. Sarlet, F. Cantrijn and D. J. Saunders: A geometrical framework for the study of non-holonomic Lagrangian systems, *J. Phys. A* **28** (1995), 3253–3268.
[32] W. Sarlet, F. Cantrijn and D. J. Saunders: A geometrical framework for the study of non-holonomic Lagrangian systems: II, *J. Phys. A* **29** (1996), 4265–4274.
[33] C. L. Siegel and J. K. Moser: *Lectures on celestial mechanics*, Springer, 1995.
[34] J. Śniatycki: Nonholonomic Noether theorem and reduction of symmetries, *Rep. Math. Phys.* **42** (1998), 5–23.
[35] J. Śniatycki: Almost Poisson spaces and nonholonomic singular reduction, *Rep. Math. Phys.* **48** (2001), 235–248.
[36] J. Śniatycki: The momentum equation and the second order differential equation condition, *Rep. Math. Phys.* **49** (2002), 371–394.
[37] H. J. Sussmann: Orbits of family of vector fields and integrability of distributions, *Trans. Amer. Math. Soc.* **180** (1973), 171–188.
[38] A. J. van der Schaft and B. M. Maschke: On the Hamiltonian formulation of nonholonomic mechanical systems, *Rep. Math. Phys.* **34** (1994), 225–233.
[39] F. W. Warner: *Foundations of differentiable manifolds and Lie groups*, Spinger, 1983.