# The inverse of a singular linear transformation and some applications in constrained mechanics



2006 J. Phys. A: Math. Gen. 39 10755

(http://iopscience.iop.org/0305-4470/39/34/011)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **39** (2006) 10755–10771 doi:10.1088/0305-4470/39/34/011

# The inverse of a singular linear transformation and some applications in constrained mechanics

**José F Cariñena**[1] **and José Fernández-Núñez**[2]

[1] Departamento de Física Teórica, Universidad de Zaragoza, 50009, Zaragoza, Spain
[2] Departamento de Física, Universidad de Oviedo, 33007 Oviedo, Spain

E-mail: jfc@unizar.es and nonius@uniovi.es

Received 6 May 2006, in final form 9 July 2006
Published 9 August 2006
Online at stacks.iop.org/JPhysA/39/10755

**Abstract**
We extend the notion of the Moore–Penrose inverse of a matrix to the case of a morphism between vector bundles and a more general concept is used to develop the geometric theory of Udwadia and Kalaba's approach to constrained Lagrangian mechanics.

PACS numbers: 02.40.−k, 45.20.Jj, 45.30.+s
Mathematics Subject Classification: 70G45, 37J05, 37J60

## 1. Introduction

The equations of motion describing the dynamical evolution of a classical mechanical system can be derived from several alternative basic principles, Newtonian, Lagrangian and Hamiltonian dynamics being the best-known examples. But there are other very useful principles and among them the so-called Gauss principle of least constraint, established by Gauss in 1829 as a consequence of the principle of virtual works, plays a relevant role.

In the process of geometrization of physics, during recent years many geometric structures have been identified as fundamental ingredients of the theory and the more detailed analysis of them from this new perspective has been very clarifying and useful, providing us with answers for different problems and proposing new questions, for instance about the uniqueness of such structures for a specific problem and the implications of the possible existence of alternative structures [4, 10, 12].

In many recent publications [15, 17, 18], and in a book [16] Udwadia and Kalaba used the Moore–Penrose inverse of a matrix to establish the equations of the motion of a system subject to non-holonomic constraints. Let us remark that the Moore–Penrose inverse has widely been used in some branches of applied mathematics, but it has been scarcely used in physics; to the best of our knowledge, it was used for the first time in mechanics in the above-mentioned paper by Udwadia and Kalaba.

Our primary aim in this paper is to derive in a geometric way the possible equations of motion of constrained systems using a generalization of the Moore–Penrose inverse of a matrix (i.e., a linear transformation) and some of its applications in constrained Lagrangian mechanics. In doing so we extend the range of application and a considerable increasing of the level of geometrical insight is obtained.

The first problem in constrained mechanics is to find consistent equations for the accelerations as functions of positions and velocities, from which to derive the trajectories fulfilling the constraints and produced by the active forces. Udwadia and Kalaba obtained the most general form of the accelerations by applying the Moore–Penrose inverse to solve the constraints when written in terms of the accelerations, for the accelerations themselves. Then they showed that the constraint force is the sum of an 'ideal' (or of D'Alembert type) term, which is uniquely determined by the unconstrained or released motion and the constraints, and another non-ideal one (i.e. a term that does work in virtual displacements), which represents influences such as friction and needs to be further specified as a defining part of the mechanical system. This information has to be included in a modified D'Alembert Principle [17]. For systems wearing only ideal constraints, both the existence and form of the Lagrange multipliers arise as a natural consequence of constructions based on fundamental principles. In this case, it is easily shown that the 'deviation' of the constrained motion from the unconstrained one is equivalent to the Gauss principle of least constraint. The extension of this principle to non-ideal constraints was given in [11].

Udwadia and Kalaba's treatment has two main restrictions: on one side, it is of a local character, sometimes it requires the use of Cartesian coordinates, and on the other side it only deals with natural Lagrangian systems, i.e. systems for which the Lagrangian is of the form $L = T - V$. Nevertheless, we think that the method worked by these authors is really useful in understanding some constructions usually made in the analysis of the dynamics of constrained mechanical systems, and it deserves some more attention in order to clarify it and to overcome such restrictions. So it is our purpose to give in this paper a detailed and geometric description of Udwadia and Kalaba approach to constrained Lagrangian mechanics.

The structure of this paper is as follows: we first give in section 2 a summary of important properties of linear algebra and a geometric construction of the inverse of a linear singular transformation between spaces more general than $\mathbb{R}^n$, both at the algebraic (section 3) and the geometric (section 4) levels. In section 5, the geometric theory of Lagrange equations when constraints are present is reviewed and the dynamical problem is formulated in terms of the underlying affine structures. Then we will apply it to the geometric analysis of constrained mechanics following the Udwadia and Kalaba approach; in particular, when only ideal constraints are present we will clarify the origin, meaning and rôle of the Lagrange multipliers (section 6) and the Gauss principle of least constraint is also studied in this context (section 7). Finally, some illustrative examples are given in section 8.

## 2. A short review of some properties of linear maps

To start with we review some basic properties of linear algebra concerning the canonical decomposition of a linear map.

Given a map $F$ between two sets $F : M \to \bar{M}$ we can define an associated equivalence relation $R_F$ in $M$ by

$$m_1 R_F m_2 \quad \Longleftrightarrow \quad F(m_1) = F(m_2),$$

in such a way that the map $F$ factorizes according to the following diagram:

$$\begin{tikzcd} M \arrow[r, "F"] \arrow[d, "\pi"'] & \bar{M} \\ M/R_F \arrow[r, "\bar{F}"'] & F(M) \arrow[u, "i"'] \end{tikzcd} \tag{1}$$

where $\pi$ denotes the canonical projection, $i$ is the natural injection and $\bar{F}$ is a bijection given by $\bar{F}([m]) = F(m)$. In the particular case in which $F$ is a bijection, $M/R_F = M$ and $F(M) = \bar{M}$ with $i = \mathrm{id}_{\bar{M}}$. We will be interested in the case of $M$ and $\bar{M}$ being linear spaces $V$ and $\bar{V}$, while $F$ will be a morphism.

We also recall that given a linear space $V$, any linear subspace $W$ has associated another equivalence relation $R_W$ by means of

$$v_1 R_W v_2 \iff v_1 - v_2 \in W.$$

When $F : V \to \bar{V}$ is a linear map between real linear spaces, the equivalence relation $R_F$ coincides with that associated with the linear subspace $\ker F$, and then $F$ can be factorized as a product of linear maps as indicated in the following diagram (see (1)):

$$\begin{tikzcd} V \arrow[r, "F"] \arrow[d, "\pi"'] & \bar{V} \\ V/\ker F \arrow[r, "\simeq", "\bar{F}"'] & F(V) \arrow[u, "i"'] \end{tikzcd} \tag{2}$$

where $\pi$ is an epimorphism, the canonical projection, $i$ is a monomorphism, the natural injection, and $\bar{F}$ is an isomorphism of linear spaces given by $\bar{F}(v + \ker F) = F(v)$.

When the linear space $V$ is a direct sum of two linear subspaces $V = W_1 \oplus W_2$, there exist projections $P_1$ and $P_2$ onto such subspaces. Recall that $P_1 + P_2 = \mathrm{id}_V$, $P_1^2 = P_1$ and $P_2^2 = P_2$. Furthermore,

$$\mathrm{Im}\, P_1 = W_1 = \ker P_2, \qquad \mathrm{Im}\, P_2 = W_2 = \ker P_1.$$

The corresponding canonical decompositions (2) of $P_1$ and $P_2$ imply that each linear subspace $W_2$ that is supplementary of the linear subspace $W_1$ is isomorphic to $V/W_1$, because $\bar{P}_2$ is an isomorphism.

When, furthermore, $V$ is endowed with an Euclidean structure, $g$, each linear subspace $W$ determines a unique orthogonal supplementary subspace

$$W^\perp = \{v \in V \mid g(v, w) = 0, \forall w \in W\},$$

and consequently an orthogonal projection $P$ onto $W$ (parallel to $W^\perp$). In this case $P = P^t$, where $P^t$ is the adjoint map of $P$. Moreover, as $W^\perp$ is a supplementary subspace of $W$ onto which $\mathrm{id}_V - P$ projects, $\overline{\mathrm{id}_V - P}$ defines an isomorphism between $V/W$ and $W^\perp$.

It is also to be remarked that the Euclidean structure of $V$ provides us with an identification of $V$ with its dual space $V^*$, by means of the linear isomorphism $\hat{g} : V \to V^*$ given by $\hat{g}(v) = g(v, \cdot)$, i.e. $\langle \hat{g}(v_1), v_2\rangle = g(v_1, v_2)$, where $\langle\cdot,\cdot\rangle$ denotes the natural pairing of the linear space $V$ with its dual. For each linear subspace $W$,

$$\hat{g}(W) = (W^\perp)^0, \qquad \hat{g}(W^\perp) = W^0,$$

where

$$W^0 = \{\alpha \in V^* \mid \langle \alpha, w\rangle = 0, \forall w \in W\}.$$

Let us now consider a linear map $F : V \to \bar{V}$ between finite-dimensional vector spaces and the linear equation $F(v) = \bar{v}$, where $\bar{v}$ is a vector $\bar{v} \in \bar{V}$; we want to determine under what conditions such an equation has a solution vector $v \in V$, i.e. there exists a vector $v \in V$ such that $F(v) = \bar{v}$, and how to determine such vectors.

The answer is given in the following theorem [8], where $F^*$ denotes the dual map of $F$, namely, the linear map $F^* : \bar{V}^* \to V^*$ given by $\langle F^*(\bar{\alpha}), v\rangle = \langle\bar{\alpha}, F(v)\rangle$.

**Theorem 1.** *Given a linear map $F : V \to \bar{V}$ between finite-dimensional vector spaces and a vector $\bar{v} \in \bar{V}$, the necessary and sufficient condition for the existence of a vector $v \in V$ such that $F(v) = \bar{v}$ is that $\langle\bar{\alpha}, \bar{v}\rangle = 0$, for all $\bar{\alpha} \in \ker F^*$. The set of such solution vectors is an affine space modelled with* $\ker F$.

Note that if $\dim V = n$ then $\operatorname{Im} F = (\ker F^*)^0$, because if $\bar{v} = F(v)$ with $v \in V$ and $\bar{\alpha} \in \ker F^*$, $\langle\bar{\alpha}, \bar{v}\rangle = \langle\bar{\alpha}, F(v)\rangle = \langle F^*(\bar{\alpha}), v\rangle = 0$, i.e. $\operatorname{Im} F \subset (\ker F^*)^0$, and when $V$ is finite dimensional both subspaces are of the same dimension. Furthermore, it is obvious that if $v - v_0 \in \ker F$, $F(v) = F(v_0)$ and conversely, if $F(v) = F(v_0)$, then $v - v_0 \in \ker F$.

## 3. Moore–Penrose inverse of a linear transformation

The concept of inverse of a singular linear transformation independently proposed by Moore and Penrose [2] can now be used. Recall that if $F : V \to \bar{V}$ is a linear map of an Euclidean space $(V, g)$, then if $P$ denotes the orthogonal projection onto the linear subspace $W = \ker F$, $\overline{\mathrm{id}_V - P}$ is an isomorphism of $V/\ker F$ in $(\ker F)^\perp$. But $\bar{F}$ is an isomorphism between $V/\ker F$ and $F(V)$, and therefore there exists an isomorphism $\sigma : \operatorname{Im} F \to (\ker F)^\perp$. In such isomorphism, if $\bar{v} \in \operatorname{Im} F$, then $\sigma(\bar{v})$ is the unique vector $v \in (\ker F)^\perp$ such that $F(v) = \bar{v}$, that is, $F\sigma(\bar{v}) = \bar{v}$; and if $v \in (\ker F)^\perp$, then $\sigma F(v) = v$. The following commutative diagram clearly shows the construction we have just made:

$$\begin{tikzcd}
V \arrow[r, "F"] \arrow[d, "\pi"'] & \bar{V} \\
V/\ker F \arrow[r, "\simeq", "\bar{F}"'] \arrow[d, "\overline{\mathrm{id}_V - P}"', "\simeq"] & F(V) \arrow[u, "i"'] \arrow[dl, "\sigma"] \\
(\ker F)^\perp &
\end{tikzcd} \tag{3}$$

Let assume that $\bar{V}$ is also endowed with an Euclidean structure $\bar{g}$. Then we give the following definition based on the diagram (3):

**Definition 1.** *The Moore–Penrose inverse of a singular linear transformation $F : V \to \bar{V}$ between two Euclidean spaces $(V, g)$ and $(\bar{V}, \bar{g})$ is the singular transformation $F^\dagger : \bar{V} \to V$ defined by*

$$F^\dagger(\bar{v}) = \begin{cases} 0 & \text{if} \quad \bar{v} \in (\operatorname{Im} F)^\perp, \\ \sigma(\bar{v}) & \text{if} \quad \bar{v} \in \operatorname{Im} F. \end{cases}$$

Note that using a similar definition for a regular linear map $F$, i.e. if $\ker F = 0$ and $\operatorname{Im} F = \bar{V}$, we would obtain the usual inverse map $F^{-1}$; therefore, the Moore–Penrose inverse is a generalization of the concept of inverse of a regular map and so we call it '(generalized) inverse'. Note that by definition $\operatorname{Im} F^\dagger = (\ker F)^\perp$, and it is also well known from linear algebra (e.g. [8]) that $\mathrm{Im}\, F^t = (\ker F)^\perp$, and therefore $\mathrm{Im}\, F^\dagger = \mathrm{Im}\, F^t$. Furthermore, $\ker F^\dagger = (\mathrm{Im}\, F)^\perp = \ker F^t$. Putting $F^t$ instead of $F$ in the last relation, we conclude that $\ker F = \ker F^{t\dagger}$.

**Proposition 1** (Moore)**.** *The linear map $F^\dagger : \bar{V} \to V$ is the Moore–Penrose inverse of $F : V \to \bar{V}$ if, and only if, $FF^\dagger$ is the orthogonal projection onto* $\mathrm{Im}\, F$ *and $F^\dagger F$ is the orthogonal projection onto* $(\ker F)^\perp$.

In fact, the map $FF^\dagger$ is given by

$$FF^\dagger(\bar{v}) = \begin{cases} 0 & \text{if} \quad \bar{v} \in (\mathrm{Im}\, F)^\perp, \\ F(\sigma(\bar{v})) = \bar{v} & \text{if} \quad \bar{v} \in \mathrm{Im}\, F. \end{cases}$$

In a similar way,

$$F^\dagger F(v) = \begin{cases} 0 & \text{if} \quad v \in \ker F, \\ \sigma F(v) = v & \text{if} \quad v \in (\ker F)^\perp. \end{cases}$$

There is an alternative characterization of $F^\dagger$, due to Penrose [14].

**Proposition 2** (Penrose)**.** *The Moore–Penrose inverse of a linear map $F : V \to \bar{V}$ is the only linear map $F^\dagger : \bar{V} \to V$ satisfying the following properties: $FF^\dagger F = F$, $F^\dagger F F^\dagger = F^\dagger$, $(FF^\dagger)^t = FF^\dagger$ and $(F^\dagger F)^t = F^\dagger F$.*

It is clear that $F^\dagger$ satisfies the above-mentioned properties as a consequence of definition 1 and proposition 1. On the other hand, if a linear map $F_1 : \bar{V} \to V$ satisfies such conditions then, $P = FF_1$ and $Q = F_1 F$ are self-adjoint projectors onto $\mathrm{Im}\, P$ and $\mathrm{Im}\, Q$, respectively; it also follows from these properties that $FF_1 = FF^\dagger$ and $F_1 F = F^\dagger F$, and consequently, $\mathrm{Im}\, P = \mathrm{Im}\, F$ and $\mathrm{Im}\, Q = (\ker F)^\perp$. The result of previous proposition shows that $F_1 = F^\dagger$.

*Linear equations in Euclidean spaces.* Let us now consider the same linear equation $F(v) = \bar{v}$ when the corresponding linear spaces are Euclidean ones. It follows from proposition 1 and theorem 1 that

**Theorem 2.** *Let $F : V \to \bar{V}$ be a linear map between Euclidean spaces. The necessary and sufficient condition for the linear equation $F(v) = \bar{v}$, $\bar{v} \in \bar{V}$, to have a solution is that $FF^\dagger(\bar{v}) = \bar{v}$. In this case $v_0 = F^\dagger(\bar{v})$ is the solution of minimal length and any other solution can be written as $v = F^\dagger(\bar{v}) + \ker F$.*

Obviously, for an arbitrary solution $v = v_0 + \xi, \xi \in \ker F$, the relation $g(v, v) \geqslant g(v_0, v_0)$ holds. We can consider in the same way the case of $g$ being negative, and thus the particular solution $v_0$ should be of maximal length.

The result in theorem 2 is still valid in the case of an inhomogeneous linear equation between affine spaces. Let $F : \mathcal{E} \to \bar{\mathcal{E}}$ be an affine map between the affine spaces $\mathcal{E}$ and $\bar{\mathcal{E}}$, which are modelled on the vector spaces $E$ and $\bar{E}$, respectively. Then, it is clear that the equation $F(x) = \bar{x}$, with a given $\bar{x} \in \bar{\mathcal{E}}$, has a solution when the associated (inhomogeneous) linear equation $[F](x - x_0) = \bar{x} - F(x_0)$ has a solution, namely, if and only if $[F][F]^\dagger(\bar{x} - F(x_0)) = \bar{x} - F(x_0)$, where $[F] \in \mathcal{L}(E, \bar{E})$ is the linear part of $F$ and $x_0 \in \mathcal{E}$ is an arbitrary origin in $\mathcal{E}$. Using the result of theorem 2, the solution of the equation $F(x) = \bar{x}$ will be expressed in the form

$$x = x_0 + [F]^\dagger(\bar{x} - F(x_0)) + \ker[F]. \tag{4}$$

*Other properties of the inverse transformation.* It is to be expected that the inverse of the inverse will be the original map, as it actually happens. Furthermore, the inverse of the adjoint is the adjoint of the inverse transformation. Note however that, in the general case, the inverse of a product is not the product of the inverses in the opposite order. In the following proposition some useful properties are pointed out.

**Proposition 3.** *The inverse $F^\dagger$ of the linear transformation $F : V \to \bar{V}$ satisfies the following properties:*

*(1)* $F^{\dagger\dagger} = F$.
*(2)* $(\alpha F)^\dagger = (1/\alpha)F^\dagger, \qquad 0 \neq \alpha \in \mathbb{R}$.
*(3)* $(F^t)^\dagger = (F^\dagger)^t$.
*(4)* $(F^t F)^\dagger = F^\dagger F^{t\dagger}$.
*(5)* $F^\dagger = F^t(FF^t)^\dagger = (F^t F)^\dagger F^t$.

Note that if $V$ and $\bar{V}$ are finite-dimensional spaces, the transpose $F^t$ and the dual transformation $F^*$ are related by $F^t = \hat{g}^{-1} F^* \hat{\bar{g}}$, and consequently the last three properties can be expressed in terms of the dual map.

As a corollary of the preceding proposition, if the linear map $F$ is injective (respectively, epijective), then $F^\dagger = (F^t F)^{-1} F^t$ (respectively, $F^\dagger = F^t(FF^t)^{-1}$), because $F^t F$ (respectively, $FF^t$) is an automorphism of $V$ (respectively, of $\bar{V}$).

## 4. Inverse morphism of a morphism of vector bundles

The previous constructions can be extended to the framework of vector bundles. Let us consider two vector bundles $\pi : E \to M$ and $\bar{\pi} : \bar{E} \to M$. The most interesting cases are when such bundles are the tangent bundle $TM$ and the cotangent bundle $T^*M$ of a differentiable manifold.

As is well known, if $F : E \to \bar{E}$ is a constant rank vector bundle morphism on the identity in the base, then $\ker F$ defines a vector subbundle of $E$ and $F(E)$ is a vector subbundle of $\bar{E}$; moreover, there exists an isomorphism $\bar{F}$ between $E/\ker F$ and $F(E)$ given by $\bar{F}(e + \ker F_{p(e)}) = F(e)$.

When the vector bundle $E$ is a direct sum of two vector subbundles, $E = E_1 \oplus E_2$, then there exist two projections $P_1$ and $P_2$ onto the respective subbundle such that

$$\operatorname{Im} P_1 = E_1 = \ker P_2, \qquad \operatorname{Im} P_2 = E_2 = \ker P_1,$$

and then we see that there exist vector bundle isomorphisms $\bar{P}_1 : E/\ker P_1 \to E_1$ and $\bar{P}_2 : E/\ker P_2 \to E_2$. The morphism $F : E \to \bar{E}$ allows us to choose a subbundle $E_2$ supplementary of $E_1 = \ker F$ in $E$ and each such a choice provides us with an isomorphism of $E_2$ with $F(E)$.

When furthermore the vector bundle $E$ is endowed with an Euclidean structure $g$, i.e. a section of the bundle $\mathcal{S}_0^2(TE) \to E$ of definite positive symmetric bilinear maps, each vector subbundle $E_1$ has a special supplementary, namely the orthogonal subbundle $E_2 = E_1^\perp$ which is determined by the Euclidean structure. Consequently, $F$ determines an isomorphism of vector bundles between $(\ker F)^\perp$ and $F(E)$.

Finally, when $\bar{E}$ is also endowed with an Euclidean structure $\bar{g}$, we can define the *generalized inverse morphism (or of Moore–Penrose)* of a constant rank morphism $F$ over the identity in the base as the linear map associating with the vectors of $(F(E))^\perp$ the null vector 0, while the image of $y \in F(E)$ is the unique vector in $(\ker F)^\perp$ whose image under $F$ is the own $y$. This inverse morphism of $F$ is denoted by $F^\dagger$ and is a linear morphism of the vector bundle $\bar{\pi}$ onto the vector bundle $\pi$.

These vector bundle maps have properties analogous to the above mentioned ones in the algebraic case. Here $P_{(\ker F)^\perp}$ and $P_{\mathrm{Im}\,F}$ are the morphisms of $\pi$ and $\bar{\pi}$, respectively, of orthogonal projection onto $(\ker F)^\perp$ and $\mathrm{Im}\,F$, respectively. Thus we have

**Proposition 4.** *The inverse morphism of a given morphism F of the vector bundle $\pi$ onto the vector bundle $\bar{\pi}$ is the only vector bundle morphism $F^\dagger$ of $\bar{\pi}$ over $\pi$ such that $P_{(\ker F)^\perp} = F^\dagger F$ and $P_{\mathrm{Im}\,F} = FF^\dagger$.*

In a similar way and with the obvious notation, the result in proposition 2 is valid at the geometric level. As the dual morphism $F^*$ and the adjoint $F^t$ of $F$ between vector bundles are well defined, the results in proposition 3 also hold.

Consider now linear equations in vector or affine bundles. Let $\pi : E \to M$ and $\bar{\pi} : \bar{E} \to M$ be two vector bundles over the manifold $M$ and let $F : E \to \bar{E}$ be a constant rank morphism of vector bundles over $\mathrm{id}_M$. A *linear equation* is a pair $(F, \bar{\sigma})$ constituted by a morphism $F$ and a section $\bar{\sigma} \in \mathrm{Sec}(\bar{\pi})$; the section $\sigma \in \mathrm{Sec}(\pi)$ is a solution of the linear equation $(F, \bar{\sigma})$ if $F \circ \sigma = \bar{\sigma}$. The space of solutions of the linear equation $(F, \bar{\sigma})$ is an affine space modelled by the vector space $\mathrm{Sec}(\pi_{|\ker F})$.

If the vector bundles are endowed with an Euclidean structure, a slight generalization of theorem 2 says us that the necessary and sufficient condition for the linear equation $(F, \bar{\sigma})$ to have a solution is that $FF^\dagger\bar{\sigma} = \bar{\sigma}$; the solutions are then expressed in the form

$$\sigma = F^\dagger \circ \bar{\sigma} + \xi, \qquad \xi \in \mathrm{Sec}(\pi_{|\ker F}). \tag{5}$$

In coordinates adapted to the vector bundle structure, a linear equation between vector bundles is expressed as a set of linear maps between fibres. Let $(q^i)$ be a system of local coordinates in a neighbourhood $U$ of the differentiable manifold $M$, and let $(q^i, x^a)$ and $(q^i, y^\alpha)$ be systems of fibre coordinates in $E$ and $\bar{E}$, respectively. The linear morphism $F$ and the section $\bar{\sigma}$ are locally expressed in the form

$$F(q^i, x^a) = \left(q^i, y^\alpha = A^\alpha_b(q)x^b\right), \qquad \bar{\sigma}(q^i) = (q^i, \bar{\gamma}^\alpha(q)),$$

in such a way that the section $\sigma(q) = (q^i, \gamma^a(q))$ solution of the linear equation $(F, \bar{\sigma})$ must satisfy in each point $q = (q^i) \in U$, the linear equation

$$A^\alpha_b(q)\gamma^b(q) = \bar{\gamma}^\alpha(q).$$

A section $\xi \in \mathrm{Sec}(\pi_{|\ker F})$ can be locally expressed as $\xi(q) = (q^i, \xi^a(q))$, together with the condition $A^\alpha_a(q)\xi^a(q) = 0$, and then the general solution (5) of the linear equation is of the following form

$$\gamma^a(q) = (A^\dagger)^a_\alpha(q)\bar{\gamma}^\alpha(q) + \xi^a(q).$$

As in the previous section, we can extend this notion and talk about a linear equation $(F, \bar{\sigma})$ in affine bundles $\mathcal{E}$ and $\bar{\mathcal{E}}$, where the morphism $F$ is affine. The linear equation has a solution iff $[F] \circ [F]^\dagger \circ (\bar{\sigma} - F \circ \sigma_0) = \bar{\sigma} - F \circ \sigma_0$ and the general solution is (4)

$$\sigma = \sigma_0 + [F]^\dagger \circ (\bar{\sigma} - F \circ \sigma_0) + \ker[F], \tag{6}$$

where $\sigma_0$ is a section of $\pi$ acting as a reference section.

## 5. Constrained Lagrangian systems

First of all, we will review the geometrical setting of the Lagrangian dynamics and explain the subsequent notation. Let $M$ be the configuration space of an $m$-dimensional autonomous Lagrangian system, and $\tau_M : TM \to M$ the tangent bundle projection. The basic constructions we will make use from now on are those of jet bundles of curves of $M$. Lagrangian systems of the first-order are described by a single function $L \in C^\infty(TM)$ (the 'Lagrangian') encoding the dynamical properties (kinematics, masses, active forces, etc) of the system; as the dynamical equations are of the second-order type, the spaces we need are the 1-jet bundle, which coincides with the tangent bundle $TM$, and the 2-jet bundle $T^2M$; both $TM$ and $T^2M$ are bundles over $M$, and the elements of $TM$ are known as 'velocities' or '(dynamical) states', while those of $T^2M$ are the 'accelerations'. There is a natural projection $\mu : T^2M \to TM$, $j^2_x\sigma \mapsto j^1_x\sigma$, $j^k_x\sigma$ denoting the $k$-jet, $k = 1, 2$, of the curve $\sigma : \mathbb{R} \to M$ at the point $\sigma(0) = x$. Every curve $\sigma$ can be prolonged to curves $j^k\sigma : \mathbb{R} \to T^kM$ given by $j^k\sigma(t) = j^k_{\sigma(t)}\sigma_t$, $\sigma_t$ being the curve $\sigma_t(s) = \sigma(t + s)$ which starts from the point $\sigma(t)$.

There exists a remarkable structure given by the 'total time derivative' operators $\mathbf{T}^0 = \mathrm{id}_{TM}$ and $\mathbf{T} : T^2M \to TTM$ defined by the rule

$$\mathbf{T} \circ j^2\sigma = (j^1\sigma)_* \circ \frac{\mathrm{d}}{\mathrm{d}t},$$

where $\mathrm{d}/\mathrm{d}t \in \mathfrak{X}(\mathbb{R})$ is the unique vector field in $\mathbb{R}$, which is endowed with a volume form $\mathrm{d}t$ (the time measure), such that $\mathrm{i}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\mathrm{d}t = 1$. These operators $\mathbf{T}^0$ and $\mathbf{T}$ are in fact vector fields along the projections $\tau_M$ and $\mu$, respectively, and are $(\mu, \tau_M)$-related, i.e. $\tau_{M*} \circ \mathbf{T} = \mathbf{T}^0 \circ \mu$. As is well known [5], each of these operators induces two derivations, one of type $d_*$ and the other of type $i_*$, along the corresponding maps; as usually, the derivations will be denoted by $d_{\mathbf{T}}$ and $i_{\mathbf{T}}$, and same for $\mathbf{T}^0$. In particular, given a function $f \in C^\infty(M)$ (resp., $f \in C^\infty(TM)$), its total time derivative is a function $d_{\mathbf{T}^0} f \in C^\infty(TM)$ (resp., $d_{\mathbf{T}} f \in C^\infty(T^2M)$).

Let us have a look at these constructions in local coordinates. Let $(q^i)$ be a local system of coordinates in $M$; let $(q^i, v^i)$ and $(q^i, v^i, a^i)$ be the corresponding fibred coordinates in $TM$ and $T^2M$. If the curve $\sigma$ locally is expressed as $\sigma(t) = (\sigma^i(t))$, $\sigma(0) = x$, the $k$-jets are $j^k_x\sigma = (\sigma^i(t), \mathrm{d}\sigma^i/\mathrm{d}t, \ldots, d^k\sigma^i/\mathrm{d}t^k)$. On the other hand, the total time derivative operators are

$$\mathbf{T}^0(q, v) = v^i \frac{\partial}{\partial q^i}, \qquad \mathbf{T}(q, v, a) = v^i \frac{\partial}{\partial q^i} + a^i \frac{\partial}{\partial v^i},$$

while the total time derivative of a function $f \in C^\infty(M)$ (resp., $f \in C^\infty(TM)$) is $d_{\mathbf{T}^0} f(q, v) = v^i \partial f/\partial q^i$ (resp., $d_{\mathbf{T}} f(q, v, a) = a^i \partial f/\partial v^i + v^i \partial f/\partial q^i$).

For our treatment of constrained dynamics, the affine structure of the fibre bundle $\mu$ is of great importance: it is an affine bundle over the vector bundle $\ker \tau_{M*}$ (also denoted by $V(\tau_M)$) of the $\tau_M$-vertical tangent vectors of $TM$. The $\tau_M$-vertical vector which corresponds to a couple of accelerations $w, w' \in \mu^{-1}(z)$ is the vector $\mathbf{T}(w) - \mathbf{T}(w') \in V_z(\tau_M)$; it will be denoted by $w - w'$. Using the general fact that a function of the total space $A$ of a fibre bundle $\pi : A \to M$ is equivalent to a fibre bundle morphism of $\pi$ over the trivial vector bundle $M \times \mathbb{R} \to M$, the total time derivative of a function $f \in C^\infty(TM)$ can be represented as a fibre bundle morphism:

$$\begin{tikzcd} T^2M \arrow[rr, "d_{\mathbf{T}} f"] \arrow[dr, "\mu"'] & & TM \times \mathbb{R} \arrow[dl, "\mathrm{pr}_1"] \\ & TM & \end{tikzcd}$$

Moreover, $d_{\mathbf{T}}f$ is an affine morphism whose linear part $[d_{\mathbf{T}}f] : V(\tau_M) \to TM \times \mathbb{R}$ is nothing but the vertical differential of $f$, i.e. the differential $\mathrm{d}f$ restricted to the subbundle $V(\tau_M)$: in fact, for two accelerations $w, w'$ over one and the same state $z \in TM$ we have

$$d_{\mathbf{T}}f(w) - d_{\mathbf{T}}f(w') = \langle \mathrm{d}f(z), \mathbf{T}(w) - \mathbf{T}(w')\rangle = \langle \mathrm{d}f(z), w - w'\rangle.$$

In local coordinates essentially we have $[d_{\mathbf{T}}f] \cdot \left(X^j \frac{\partial}{\partial v^j}\right) = X^j \frac{\partial f}{\partial v^j}$.

*Fibre derivatives; the Euler–Lagrange form and the Hessian.* The concept of fibre derivative of a morphism of affine bundles [9] is another useful geometric tool in our constructions. Given a function $f \in C^\infty(TM)$, its *fibre derivative* is a morphism $\mathcal{F}f : TM \to T^*M$ given by $\mathcal{F}f(z) = Df_{\tau_M(z)}(z)$, where $f_x$ is the restriction of $f$ to $T_xM$ and symbol $D$ stands for the derivative of an application between linear spaces. The second fibre derivative is a linear morphism $\mathcal{F}^2 f : TM \to \mathcal{S}^2(TM)$ from $TM$ to the bilinear symmetric functions of $TM$ given by $\mathcal{F}^2 f(z) = D^2 f_{\tau_M(z)}(z)$. By means of the vertical lift $\xi^{\mathrm{v}} : TM \times_M TM \to V(\tau_M)$ [3], we extend this morphism to a bundle morphism $\mathcal{H}f : TM \to \mathcal{S}^2(V(\tau_M))$ according to the rule

$$\mathcal{H}f(z) \cdot (\xi^{\mathrm{v}}(z, v), \xi^{\mathrm{v}}(z, v')) = \mathcal{F}^2 f(z) \cdot (v, v'), \qquad v, v' \in T_{\tau_M(z)}M.$$

In this sense, the second fibre derivative of a function $f$ is usually called the *Hessian map of* $f$. The induced linear morphism between the vector bundle $V(\tau_M)$ and its dual vector bundle $V^*(\tau_M)$ (the bundle of $\tau_M$-semibasic 1-forms), will be denoted by $\widehat{\mathcal{H}f}$ and the following diagram holds

$$\begin{tikzcd} V(\tau_M) \arrow[rr, "\widehat{\mathcal{H}f}"] \arrow[dr] & & V^*(\tau_M) \arrow[dl] \\ & TM & \end{tikzcd} \tag{7}$$

This is the interpretation of the second fibre derivative we need in the following, and it can be proved [9] that $\ker(\mathcal{F}f)_* = \ker \widehat{\mathcal{H}f}$, a property saying that $\mathcal{F}f$ is a (local) diffeomorphism at a point $z \in TM$ if and only if $\widehat{\mathcal{H}f}_{|V_z(\tau_M)}$ is a linear isomorphism. When this condition holds, the function $f$ is said to be *regular* (or *hyper-regular*, for a global diffeomorphism) and the Hessian map of $f$, non-degenerate. Consequently, the vertical bundle $V(\tau_M)$ will be endowed with a scalar product structure $\mathcal{H}f$; when, furthermore, the Hessian map is positive, the scalar product in $V(\tau_M)$ is Euclidean.

In local coordinates the expressions for the fibre derivatives are

$$\mathcal{F}f(q, v) = \frac{\partial f}{\partial v^i}\,\mathrm{d}q^i, \qquad \widehat{\mathcal{H}f} \cdot \left(X^i \frac{\partial}{\partial v^i}\right) = X^i \frac{\partial^2 f}{\partial v^i \partial v^j}\,\mathrm{d}q^j.$$

The regularity condition will be $\det(\partial^2 f/\partial v^i \partial v^j) \neq 0$.

Associated with every function $f \in C^\infty(TM)$ we define a 1-form on the space of accelerations, $T^2M$, known as the *Euler–Lagrange form* $\delta f$,

$$\delta f = d_{\mathbf{T}}\theta_f - \mu^*\,\mathrm{d}f \in \bigwedge^1(T^2M), \tag{8}$$

where $\theta_f = \mathrm{d}f \circ S \in \bigwedge^1(TM)$ is a $\tau_M$-semibasic form, i.e. $\theta_f \in \mathrm{Sec}(V^*(\tau_M))$, defined by means of the almost-tangent structure (or vertical endomorphism) $S$ of the tangent bundle $T(TM)$. Note that the 1-form along $\tau_M$ equivalent to $\theta_f$ is nothing but the fibre derivative of $f$, and also that $\delta f$ can be expressed in terms of the 'energy' $E_f = \Delta(f) - f$, where $\Delta \in \mathfrak{X}(TM)$ is the Liouville vector field [3]:

$$\delta f = i_{\mathbf{T}}\,\mathrm{d}\theta_f + \mu^*\,\mathrm{d}E_f.$$

As the Euler–Lagrange form $\delta f$ is $\tau \circ \mu$-semibasic, it is equivalent to a 1-form along $\tau_2 = \tau \circ \mu$, i.e. to a bundle morphism $\delta f^{\vee} : T^2M \to T^*M$; using again the vertical lift, we can extend $\delta f$, by duality, to an equivalent morphism $\widetilde{\delta f}$ from $T^2M$ to the space of $\tau_M$-semibasic 1-forms $V^*(\tau_M)$:

$$\begin{tikzcd} T^2M \arrow[rr, "\widetilde{\delta f}"] \arrow[dr, "\mu"'] & & V^*(\tau_M) \arrow[dl] \\ & TM & \end{tikzcd} , \qquad (9)$$

the relation between $\delta f^{\vee}$ and $\widetilde{\delta f}$ being

$$\langle \delta f^{\vee}(w), v\rangle = \langle \widetilde{\delta f}(w), \xi^{\mathrm{v}}(z, v)\rangle, \qquad w \in \mu^{-1}(z), \quad v \in T_{\tau_M(z)}M. \qquad (10)$$

It is this interpretation (9) of the Euler–Lagrange 1-form what we need later in the theory of constrained dynamics. The coordinate representation of the Euler–Lagrange form is

$$\delta f = \left[ d_{\mathbf{T}}\left(\frac{\partial f}{\partial v^j}\right) - \mu^* \frac{\partial f}{\partial q^j}\right] \mathrm{d}q^j.$$

The main property we want to point out is as follows.

**Proposition 5.** *The Euler–Lagrange 1-form $\widetilde{\delta f}$ is an affine morphism of affine bundles whose linear part is the Hessian map of $f$.*

**Proof.** Let us consider two accelerations $w, w' \in \mu^{-1}(z)$; thus $\widetilde{\delta f}(w), \widetilde{\delta f}(w') \in V_z^*(\tau_M)$ and for a vector $v \in T_zM$ we have (10) $\langle \widetilde{\delta f}(w) - \widetilde{\delta f}(w'), \xi^{\mathrm{v}}(z, v)\rangle = \langle \delta f^{\vee}(w) - \delta f^{\vee}(w'), v\rangle$. Now, choosing two vectors $Y \in T_wT^2M$ and $Y' \in T_{w'}T^2M$ such that $(\tau_M \circ \mu)_{*w} \cdot Y = (\tau_M \circ \mu)_{*w'} \cdot Y' = v$, a direct calculation, using the properties of all the differential forms involved, gives

$$\langle \widetilde{\delta f}(w) - \widetilde{\delta f}(w'), \xi^{\mathrm{v}}(z, v)\rangle = \mathrm{d}\theta_f(z) \cdot (\mathbf{T}(w) - \mathbf{T}(w'), \mu_{*w} \cdot Y). \qquad (11)$$

On the other hand, a direct calculation, this time using the (local) flow of the vertical lift $X^{\mathrm{v}} \in \mathfrak{X}(TM)$ of a vector field $X \in \mathfrak{X}(M)$, shows that

$$\begin{aligned} \mathrm{d}\theta_f(z) \cdot (X^{\mathrm{v}}(z), V) &= \langle \pounds_{X^{\mathrm{v}}}\theta_f(z), V\rangle \\ &= \mathcal{F}^2 f(z) \cdot (X(\tau_M(z)), \tau_{M*z} \cdot V) \\ &= \mathcal{H}f(z) \cdot (X^{\mathrm{v}}(z), \xi^{\mathrm{v}}(z, \tau_{M*z} \cdot V)); \end{aligned}$$

applying this result to the previous partial one (11) we finally arrive at

$$\langle \widetilde{\delta f}(w) - \widetilde{\delta f}(w'), \xi^{\mathrm{v}}(z, v)\rangle = \mathcal{H}f(z) \cdot (w - w', \xi^{\mathrm{v}}(z, v)), \qquad \forall v \in T_zM,$$

that is

$$\widetilde{\delta f}(w) - \widetilde{\delta f}(w') = \widehat{\mathcal{H}f}(w - w'), \qquad (12)$$

$\forall w, w' \in T^2M$ such that $\mu(w) = \mu(w')$. □

*Unconstrained motion.* Now let us study the unconstrained motion of a first-order Lagrangian system. There are two alternative but equivalent geometrical interpretations of the Euler–Lagrange equations of the variational problem with Lagrangian $L$ [5]: the extremals can be regarded either as the integral curves of a second-order differential equation (a SODE, for short) $\Gamma \in \mathfrak{X}(TM)$ solution of the equation $\mathrm{i}(\Gamma)\,\mathrm{d}\theta_L = -\mathrm{d}E_L$, or as those of a section $\gamma \in \mathrm{Sec}(\mu)$ such that $\gamma^*\delta L = 0$, that is $\delta L^{\vee} \circ \gamma = 0$ or

$$\widetilde{\delta L} \circ \gamma = 0; \qquad (13)$$

thus, the dynamical equation is interpreted as a linear equation in affine bundles, which is the suitable point of view in our constructions. The relation between $\gamma$ and $\Gamma$ is given by

$$\Gamma = \mathbf{T} \circ \gamma; \tag{14}$$

in local coordinates, if $\Gamma(q, v) = v^j \frac{\partial}{\partial q^j} + \Gamma^j(q, v)\frac{\partial}{\partial v^j}$, then the equivalent section reads $\gamma(q, v) = (q^j, v^j, \Gamma^j(q, v))$.

From now on, we will consider only regular Lagrangians, i.e. Lagrangians $L$ such that $\mathcal{F}L_*$ is a (local) diffeomorphism. As is well established, in such cases the solution is unique, that is, there is a unique (local) SODE $\gamma_0 \in \mathrm{Sec}(\mu)$ solution of the linear equation (13). This is the unconstrained (or *released*) motion obeying the Hamilton principle with $L$ as Lagrangian, and is completely determined by $L$.

*Constrained motion.* The main goal of this paper is to study the constrained motion of a regular Lagrangian system using the concept of inverse we have explained in the previous sections.

The first fundamental problem in constrained dynamics is to find a system of consistent differential equations whose integral curves are the trajectories which are compatible with both the applied forces and the restrictions imposed by the constraints.

Geometrically speaking, a *constraint* of the dynamical system is a submanifold $N \subset TM$ of the space of states where the evolution takes place. $N$ can be defined, at least locally, by $N = f^{-1}(0)$, where $f = (f^1, \ldots, f^r) \in C^\infty(M, \mathbb{R}^r)$ is the *constraint function*, and $r$ the codimension of $N$. The $r$ functions $f^\alpha \in C^\infty(TM)$, $\alpha = 1, \ldots, r$, impose restrictions on the states by the constraint equations $f^\alpha(z) = 0$, and are supposed to be functionally independent, although this condition is not essential in what follows.

We do not assume that the constraint function has some special features, for instance linearity in the velocities; we are considering any type of constraints usually appearing in (time-independent) mechanics. For holonomic constraints, i.e. constraints of the form $f(q) = 0$, the constraint function is not $\tau_M^* f$ but the function $d_{\mathrm{T}^0} f$; the constraint $d_{\mathrm{T}^0} f = 0$ means that during the motion the system has to satisfy the geometric constraint.

In the same sense, obvious conditions of consistence of the motion lead to impose the restrictions

$$d_{\mathbf{T}} f(w) = 0 \tag{15}$$

on the space of accelerations, where $d_{\mathbf{T}} f$ is the action of the total time derivative operator extended to $f$, i.e. $d_{\mathbf{T}} f = (d_{\mathbf{T}} f^1, \ldots) \in C^\infty(T^2M, \mathbb{R}^r)$. The admissible accelerations $w \in \mu^{-1}(z)$ over a given state $z$ are just those satisfying (15). It is important to observe that the difference of two admissible accelerations $w$, $w'$ over one and the same state $z$ is a vector $\eta \in V_z(\tau_M)$ belonging to the kernel of $[d_{\mathbf{T}} f]$; these vectors are known as *virtual displacements* (or, more properly, *virtual velocities* [1]) at $z$.

If we look for a SODE $\gamma \in \mathrm{Sec}(\mu)$ giving the admissible accelerations over the states during a motion compatible with the constraints, we have to consider the linear equation $d_{\mathbf{T}} f \circ \gamma = 0$, or equivalently

$$[d_{\mathbf{T}} f] \circ (\gamma - \gamma_0) = \mathbf{b}, \tag{16}$$

where the term $\mathbf{b} = -d_{\mathbf{T}} f \circ \gamma_0 \in \mathrm{Sec}(TM \times \mathbb{R}^r)$ is completely determined by the constraints and the unconstrained motion. Based on these considerations we give the following definitions.

**Definition 2.** *An 'admissible' or 'allowable' motion is a section $\gamma \in \mathrm{Sec}(\mu)$ such that its deviation $\gamma - \gamma_0 \in \mathrm{Sec}(V(\tau_M))$ from the released motion $\gamma_0$ obeys the linear equation (16). And a 'virtual displacement' is a $\tau_M$-vertical vector field belonging to the kernel of* $[d_{\mathbf{T}} f]$.

Thus, if $\gamma$ is an allowable motion, so is $\gamma + X$, for all virtual displacements $X \in \ker[d_{\mathbf{T}} f] \subset \mathrm{Sec}(V(\tau_M))$.

As far as the dynamical equation is concerned we could state a variational formulation taking the path variations restricted by the constraint $N$ (a *vakonomic* formulation of the Hamilton principle, see e.g. [1]), but we adopt the point of view of the D'Alembert–Lagrange principle, though we do not assume *a priori* any special nature for the constraint force. That is to say, we assume that the actual motion does not satisfy equation (13), and in consequence there should exist a nonzero $\tau_M$-semibasic 1-form $Q$, called the *constraint force*, such that the actual motions $\gamma$ are solutions of the linear (non-homogeneous) equation

$$\widetilde{\delta L} \circ \gamma = Q. \tag{17}$$

Using the affine structures involved, we can write the following linear equation for the deviation $\gamma - \gamma_0 \in \mathrm{Sec}(V(\tau_M))$ of the actual motion from the released one (which does satisfy (13)):

$$\widehat{\mathcal{H}L} \circ (\gamma - \gamma_0) = Q. \tag{18}$$

If we knew the constraint force $Q \in \mathrm{Sec}(V^*(\tau_M))$, this equation immediately would provide the solution $\gamma - \gamma_0 = \widehat{\mathcal{H}L}^{-1} Q$. Unfortunately, $Q$ is completely unknown, but although it were the case, perhaps this solution did not verify the constraint equation (16) and it would be useless. The very dynamical problem is to look for the sections $\gamma - \gamma_0 \in \mathrm{Sec}(V(\tau_M))$ satisfying the linear equations (16) and (18); we can visualize it in the following commutative diagram:

$$\begin{tikzcd}
TM \times \mathbb{R}^r \arrow[dr, shift left] & V(\tau_M) \arrow[l, "{[d_{\mathbf{T}} f]}"'] \arrow[r, "\widehat{\mathcal{H}L}"] \arrow[d, shift left] & V^*(\tau_M) \arrow[dl, shift left] \\
& TM \arrow[ul, "\mathbf{b}"] \arrow[u, "\gamma - \gamma_0"] \arrow[ur, "Q"'] &
\end{tikzcd} \tag{19}$$

In the remaining sections we give a detailed analysis of these equations, studying the nature and some properties of their solutions.

## 6. The constraint force and the Lagrange multiplier

Once the dynamical problem has been formulated, the first thing we must check is the consistence of the linear equations (16) and (18). As the Lagrangian $L$ is regular, a scalar product structure is defined in the space $V(\tau_M)$, and equation (18) is certainly consistent. On the other hand, in the space $TM \times \mathbb{R}^r$ the Euclidean structure induced by the canonical Euclidean product $g_r$ in $\mathbb{R}^r$ can be considered, and, with reference to these two scalar product structures, an inverse of the linear morphism $[d_{\mathbf{T}} f] : V(\tau_M) \to TM \times \mathbb{R}^r$ is defined. Remembering that $[d_{\mathbf{T}} f]^\dagger$ is a linear morphism from $TM \times \mathbb{R}^r$ to $V(\tau_M)$, we can state the consistence of equation (16) requiring that (theorem 2)

$$[d_{\mathbf{T}} f] \circ [d_{\mathbf{T}} f]^\dagger \circ \mathbf{b} = \mathbf{b}, \tag{20}$$

this condition represents a sort of secondary constraint and only in the set of points $N_1 \subseteq N$ where it is satisfied the dynamical problem will have solution. That is, the constraint submanifold is a smaller submanifold $N_1 \subseteq N \subset TM$. The condition (20) is automatically fulfilled when the constraint function $f$ is a submersion, i.e. the constraints $f^\alpha$ are independent.

As pointed out above, the force of constraint $Q$ is unknown, but it is possible to determine its form at every admissible state in terms of some known quantities plus additional information about the non-ideal character of constraints [15, 16].

In fact, as an immediate consequence of theorem 2, the general solution of equation (16), in the region $N_1$ where it is consistent, can be written as (6)

$$\gamma - \gamma_0 = [d_{\mathbf{T}} f]^{\dagger} \circ \mathbf{b} + \eta, \tag{21}$$

where $\eta$ is an arbitrary $[d_{\mathbf{T}} f]$-vertical section of $V(\tau_M)$, i.e. a virtual displacement. Substituting this expression in equation (18) we obtain an explicit expression giving the form of the constraint force $Q$:

$$Q = \widehat{\mathcal{H}L} \circ [d_{\mathbf{T}} f]^{\dagger} \circ \mathbf{b} + \widehat{\mathcal{H}L} \circ \eta. \tag{22}$$

Thus the force of constraint $Q$ is made up of two additive terms. The first term is explicitly determined by the constraints and the released motion, while the second one is left completely arbitrary [13, 16].

The first term in (22) enjoys the important property of being transversal to the virtual displacements. In fact, for every virtual displacement $\zeta_z$ at the state $z$, we have

$$\langle \widehat{\mathcal{H}L}([d_{\mathbf{T}} f]^{\dagger} \cdot \mathbf{b}(z)), \zeta_z \rangle = \mathcal{H}L(z) \cdot ([d_{\mathbf{T}} f]^{\dagger} \cdot \mathbf{b}(z), \zeta_z) = 0$$

because the definition of the inverse $[d_{\mathbf{T}} f]^{\dagger}$ says that $\mathrm{Im}[d_{\mathbf{T}} f]^{\dagger} = (\ker[d_{\mathbf{T}} f])^{\perp}$, with the scalar structure on $V(\tau_M)$ provided by $\mathcal{H}L$. In mechanics, the product $\langle Q(z), \zeta_z \rangle$ is usually known as the 'virtual work' of the $Q$ force at $z$, and so the property we have pointed out means that the force of constraint $Q$ has a component that does not work in any virtual displacement. That is, it obeys the D'Alembert principle (of virtual work); for that reason, we will call it *ideal* or *of D'Alembert type*, and denote it by $Q_{\mathrm{id}}$:

$$Q_{\mathrm{id}} = \widehat{\mathcal{H}L} \circ [d_{\mathbf{T}} f]^{\dagger} \circ \mathbf{b}. \tag{23}$$

The other component, denoted by $Q_{\mathrm{nid}}$, does work on the system in a virtual displacement, it is not ideal. In general, for such a constraint force

$$\langle Q_{\mathrm{nid}}(z), \zeta_z \rangle = \mathcal{H}L(z) \cdot (\eta_z, \zeta_z) \neq 0, \qquad \eta \in \mathrm{Sec}(\ker[d_{\mathbf{T}} f]). \tag{24}$$

So, in the presence of constraints we are led to the existence of a necessary constraint force $Q_{\mathrm{id}}$ that can be understood as the 'minimal' force necessary to fulfil the constraint. In order to establish unambiguous dynamical equations, the non-ideal part of the force of constraint must be specified by means of new additional (mechanical) information about the constraint, going beyond that included in the constraint function $f$ itself and representing, for instance, effects such as friction. The needed information can be provided by a prescription of the virtual work done by the constraint; i.e. we specify the scalar product $\mathcal{H}L \cdot (\eta, \zeta)$ as a known function of $\eta$ for all virtual displacements $\zeta$. The choice would be part of the mechanical model we are dealing with. This is what is given in practice when we assume, for instance, that the sliding friction is of Coulomb or of Stokes type. For more details and illustrative examples, see [13, 16, 17].

*The Lagrange multiplier.* Lagrange's multipliers are usually introduced to write consistent differential equations for the motion of constrained systems, but in the usual treatment their existence is not a consequence of basic laws of motion. However, in our formulation they arise as a simple consequence of the properties of the inverse $[d_{\mathbf{T}} f]^{\dagger}$.

For an ideal constraint, the application of proposition 3.5 in terms of the dual map $[d_{\mathbf{T}} f]^*$, with $g = \mathcal{H}L$ and $\bar{g} = g_r$, to (23) gives

$$Q_{\mathrm{id}} = [d_{\mathbf{T}} f]^* \circ \widehat{g_r} \circ ([d_{\mathbf{T}} f] \circ [d_{\mathbf{T}} f]^t)^{\dagger} \circ \mathbf{b},$$

that is,

$$Q_{\mathrm{id}} = [d_{\mathbf{T}} f]^* \circ \Lambda, \tag{25}$$

with

$$\Lambda = \widehat{g_r} \circ ([d_{\mathbf{T}} f] \circ [d_{\mathbf{T}} f]^t)^\dagger \circ \mathbf{b} \in \mathrm{Sec}(TM \times \mathbb{R}^{*r}). \tag{26}$$

$\Lambda$ is the *Lagrange multiplier*; it is a $r$-dimensional 'real' representation of the constraint force $Q$ and its components are known as the 'Lagrange multipliers'. We insist on the fact that the existence and explicit expression of $\Lambda$ derives from the basic assumptions we made.

## 7. The Gauss principle of least constraint

Let us consider the case in which only ideal constraints exist, i.e. all constraints on the system are such that the constraint forces are ideal. Once the constraint force is known, equation (21) immediately gives for the actual motion of the system the explicit expression

$$\gamma = \gamma_0 + [d_{\mathbf{T}} f]^\dagger \circ \mathbf{b}, \tag{27}$$

this solution is unique and the trajectory of the system is the integral curve of $\gamma$ satisfying a given initial condition. When the Hessian of the Lagrangian is positive, equation (27) admits a direct and nice interpretation according to theorem 2: the deviation $\gamma - \gamma_0$ is of minimal length in the Euclidean metric provided by $\mathcal{H}L$.

Defining as a measure of the deviation of motions from the released one the quadratic form (the 'Gaussian constraint')

$$\mathfrak{G}(\gamma) = \mathcal{H}L \cdot (\gamma - \gamma_0, \gamma - \gamma_0), \tag{28}$$

it is easy to show that the admissible motion minimizing the Gaussian constraint is nothing but (27). Consequently, we have proved the following theorem.

**Theorem 3** (Gauss' principle of least constraint)**.** *An admissible motion (16) is the actual motion of the system if and only if it minimizes the Gaussian constraint (28) at each state.*

Seeing the deviation $\gamma - \gamma_0$ as the vector $\widehat{\mathcal{H}L}$-equivalent of the force of constraint $Q_{\mathrm{id}}$ (see (23)), the force of constraint is also minimal.

The Gauss principle is valid at each state of velocity $z \in N_1$; it states that among all admissible accelerations $w \in \mu^{-1}(z)$ the actual one chosen by the system is the one minimizing the Gaussian constraint. This principle is equivalent to the D'Alembert–Lagrange construction followed by us, and can be put at the foundations of classical mechanics.

## 8. Examples

Let us apply these constructions to some known examples taken from classical mechanics. We will use subscripts instead of superscripts in coordinates $q_i$, velocities $v_i$ and accelerations $a_i$.

**Example 1.** Let us consider the Lagrangian $L \in C^\infty(T(\mathbb{R}^2 \times S^2))$ given in local coordinates by

$$L = \tfrac{1}{2} m\left(v_1^2 + v_2^2\right) + \tfrac{1}{2} I_3 v_3^2 + \tfrac{1}{2} I_4 v_4^2, \tag{29}$$

and subject to the non-holonomic constraints

$$v_1 - R v_4 \cos q_3 = 0, \qquad v_2 - R v_4 \sin q_3 = 0, \tag{30}$$

with $m$, $I_3$, $I_4$ and $R$ being positive constants. This is the case of a rolling-spinning disc on a horizontal plane: $m$ is the mass, $R$ the radius and $I_3$, $I_4$ are the momenta of inertia with respect to the spinning and rolling axes, respectively. The unconstrained motion is $\gamma_0(q, v) = (q, v; 0)$, and the Hessian is given by $\widehat{\mathcal{H}L}(q, v) = \mathrm{diag}(m, m, I_3, I_4)$.

The total time derivative of constraints (30) is

$$d_{\mathbf{T}}f(q,v,a)=\begin{pmatrix}a_1-Ra_4\cos q_3+Rv_3v_4\sin q_3\\ a_2-Ra_4\sin q_3-Rv_3v_4\cos q_3\end{pmatrix}\in\{(q,v)\}\times\mathbb{R}^2,$$

so its linear part $[d_{\mathbf{T}}f]:V(\tau_M)\to TM\times\mathbb{R}^2$ is represented on the basis $\{\partial/\partial v_i\}$ of $V_{(q,v)}(\tau_M)$ and the canonical basis $(\mathbf{e}_1,\mathbf{e}_2)$ of $\mathbb{R}^2$ by the matrix

$$[d_{\mathbf{T}}f](q,v)=\begin{pmatrix}1&0&0&-R\cos q_3\\ 0&1&0&-R\sin q_3\end{pmatrix},$$

and the vector $\mathbf{b}(q,v)=-d_{\mathbf{T}}f(q,v;0)$ (16) is given by

$$\mathbf{b}(q,v)=Rv_3v_4(-\sin q_3\mathbf{e}_1+\cos q_3\mathbf{e}_4)\in\{(q,v)\}\times\mathbb{R}^2.$$

In order to find the inverse of $[d_{\mathbf{T}}f]$ we have that $\ker[d_{\mathbf{T}}f](q,v)$ is generated by the vectors $R\cos q_3\mathbf{e}_1+R\sin q_3\mathbf{e}_2+\mathbf{e}_4$ and $\mathbf{e}_3$, while $(\ker[d_{\mathbf{T}}f])^{\perp}$ is generated by the vectors $\mathbf{e}_1-(mR/I_4)\cos q_3\mathbf{e}_4$ and $\mathbf{e}_2-(mR/I_4)\sin q_3\mathbf{e}_4$. As the constraint function $f$ is epijective, we also have $\mathrm{Im}[d_{\mathbf{T}}f]=\mathbb{R}^2$. Thus, the M–P inverse we are looking for is

$$[d_{\mathbf{T}}f]^{\dagger}(q,v)=\frac{1}{I_4+mR^2}\begin{pmatrix}I_4+mR^2\sin^2 q_3 & -mR^2\sin q_3\cos q_3\\ -mR^2\sin q_3\cos q_3 & I_4+mR^2\cos^2 q_3\\ 0&0\\ -\frac{mR}{I_4}\cos q_3 & -\frac{mR}{I_4}\sin q_3\end{pmatrix}.$$

Assuming that the constraints are ideal, the deviation from the released motion is $(\gamma-\gamma_0)(q,v)=[d_{\mathbf{T}}f]^{\dagger}\cdot\mathbf{b}(q,v)=Rv_3v_4(-\sin q_3\partial/\partial v_1+R\cos q_3\partial/\partial v_2)$, and gives rise to the consistent system of second-order differential equations determining all the constrained motions (cf f.i. [6])

$$\ddot{q}_1=-R\dot{q}_3\dot{q}_4\sin q_3,\qquad \ddot{q}_2=R\dot{q}_3\dot{q}_4\cos q_3,\qquad \ddot{q}_3=0,\qquad \ddot{q}_4=0.$$

To find the Lagrange multiplier $\Lambda$ and the (ideal) constraint force $Q$ we first note that $[d_{\mathbf{T}}f][d_{\mathbf{T}}f]^t$ is invertible, and then its M–P inverse is the inverse matrix $A=([d_{\mathbf{T}}f][d_{\mathbf{T}}f]^t)^{-1}$,

$$A=\frac{m}{I_4+mR^2}\begin{pmatrix}I_4+mR^2\sin^2 q_3 & -mR^2\sin q_3\cos q_3\\ -mR^2\sin q_3\cos q_3 & I_4+mR^2\cos^2 q_3\end{pmatrix},$$

consequently

$$\Lambda(q,v)=\widehat{g_2}([d_{\mathbf{T}}f][d_{\mathbf{T}}f]^t)^{\dagger}\mathbf{b}(q,v)=mRv_3v_4(-\sin q_3\varepsilon^1+\cos q_3\varepsilon^2),$$

with $(\varepsilon^1,\varepsilon^2)$ being the dual basis of $(\mathbf{e}_1,\mathbf{e}_2)$, and

$$Q(q,v)=[d_{\mathbf{T}}f]^*\Lambda(q,v)=mRv_3v_4(-\sin q_3 dv_1+\cos q_3 dv_2).$$

We could also evaluate the constraint force by using the expression $Q=\widehat{\mathcal{H}L}\circ\gamma$, and we shall obtain, of course, the same result.

**Example 2.** Sometimes it is of interest to ask for the force acting on a mechanical system in order to have a partially known motion, that is, a motion of which we know some characteristics, like the equation of the orbit or some first integrals. A typical and important inverse problem of this kind is the very well-known Kepler problem, already treated in [16] with the aid of the M–P inverse in Cartesian coordinates. Let us try to find now the force responsible for the motion of a particle of mass $m$ in an aerial-velocity preserving elliptic orbit centred at a point $O$. We know that the (free) Lagrangian $L\in C^{\infty}(T(\mathbb{R}^2-O))$ is expressed in polar coordinates $(r,\varphi)$ centred at $O$ as follows:

$$L=\tfrac{1}{2}m\left(v_r^2+r^2v_\varphi^2\right).\tag{31}$$

If we fix $\varphi = 0$ at one pericentre, the equation of the orbit and the conservation of the aerial-velocity (or Kepler's second law) are

$$\frac{p}{r^2} = 1 + \varepsilon\cos(2\varphi), \qquad r^2 v_\varphi = C, \tag{32}$$

where $p$, $\varepsilon$ and $C$ are constants (with $\varepsilon < 1$). The problem is to look for the ideal constraint force of constraints (32) which apart the Lagrangian system (31) from the released motion $a_r^0 = r v_\varphi^2$, $a_\varphi^0 = -\frac{2}{r} v_r v_\varphi$.

The total time derivative of the constraints at the point with coordinates $(r, v, v_r, v_\varphi, a_r, a_\varphi)$ is the matrix

$$d_{\mathbf{T}} f = \begin{pmatrix} \frac{2p}{r} a_r - 2\varepsilon r (r a_\varphi + 4 v_r v_\varphi)\sin 2\varphi + \frac{2p}{r^2} v_r^2 - 4\varepsilon r^2 v_\varphi^2 \cos 2\varphi \\ r^2 a_\varphi + 2 r v_r v_\varphi \end{pmatrix},$$

its linear part is represented by the regular matrix

$$[d_{\mathbf{T}} f] = \begin{pmatrix} \frac{2p}{r} & -2\varepsilon r^2 \sin 2\varphi \\ 0 & r^2 \end{pmatrix},$$

whose M–P inverse is its inverse matrix

$$[d_{\mathbf{T}} f]^\dagger = [d_{\mathbf{T}} f]^{-1} = \frac{1}{2pr} \begin{pmatrix} r^2 & 2\varepsilon r^2 \sin 2\varphi \\ 0 & \frac{2p}{r} \end{pmatrix}.$$

Taking into account constraints (32), the vector $\mathbf{b}$ is given by

$$\mathbf{b} = -\frac{2C^2(1-\varepsilon^2)}{p^2}\mathbf{e}_1,$$

and consequently the deviation $\gamma - \gamma_0$ is

$$\gamma - \gamma_0 = [d_{\mathbf{T}} f]^\dagger \mathbf{b}(r, \varphi, v_r, v_\varphi) = -\frac{C^2(1-\varepsilon^2)}{p^2} r \frac{\partial}{\partial v_r}.$$

Finally, the constraint force will be

$$Q(r, \varphi, v_r, v_\varphi) = \widehat{\mathcal{H}L} \circ \gamma(r, \varphi, v_r, v_\varphi) = -\frac{mC^2(1-\varepsilon^2)}{p^2} r\, dv_r,$$

with the Hessian $\widehat{\mathcal{H}L}(r, \varphi, v_r, v_\varphi) = \mathrm{diag}(m, mr^2)$. We recognize here an elastic force of Hooke's type centred at the point $O$.

More examples can be seen in [16].

## References


[1] Arnold V I (ed) 1988 *Dynamical Systems III* (Berlin: Springer) chapter I, section 2
[2] Ben-Israel A 2002 The Moore of the Moore–Penrose inverse *Elect. J. Linear Algebra* **9** 150–7
[3] Cariñena J F 1990 Theory of singular Lagrangians *Fortschr. Phys.* **38** 641–80
[4] Cariñena J F and Ibort L A 1983 Non-Noether constants of motion *J. Phys. A: Math. Gen.* **16** 1–7
[5] Cariñena J F, López C and Martínez E 1989 A new approach to the converse of Noether's theorem *J. Phys. A: Math. Gen.* **22** 4777–87
[6] Cariñena J F and Rañada M F 1993 Lagrangian systems with constraints: a geometrical approach to the method of Lagrange multipliers *J. Phys. A: Math. Gen.* **26** 1335–51
[7] Casey J 1994 Geometrical derivation of Lagrange's equations for a system of particles *Am. J. Phys.* **62** 836–47
[8] Chambadal L and Ovaert J L 1968 *Algèbre Linéaire et Algèbre Tensorielle* (Paris: Dunod)
[9] Gràcia X 2000 Fibre derivatives: some applications to singular Lagrangians *Rep. Math. Phys.* **45** 67–84
[10] Hojman S and Harleston H 1981 Equivalent Lagrangians: multidimensional case *J. Math. Phys.* **22** 1414–19
[11] Kalaba R, Natsuyama H and Udwadia F 2004 An extension of Gauss's principle of least constraint *Int. J. Gen. Sci.* **33** 63–9

[12] Magri F 1978 A simple model of the integrable Hamiltonian equation *J. Math. Phys.* **1** 1156–62

[13] O'Reilly O M and Srinivasa A R 2001 On a decomposition of generalised constraint forces *Proc. R. Soc.* A **457** 1307

[14] Penrose R 1955 A generalized inverse for matrices *Proc. Camb. Phil. Soc.* **51** 406–13

[15] Udwadia F E and Kalaba R E 1992 A new perspective in constrained motion *Proc. R. Soc.* A **439** 407–10

[16] Udwadia F E and Kalaba R E 1996 *Analytical Dynamics: A New Approach* (Cambridge: Cambridge University Press)

[17] Udwadia F E and Kalaba R E 2000 Nonideal constraints and Lagrangian dynamics *J. Aerosp. Eng.* **13** 17–22

[18] Udwadia F E and Kalaba R E 2002 On the foundations of analytical dynamics *Int. J. Non-Linear Mech.* **37** 1079–90