



**HAL**  
open science

# Lagrange Multipliers in Infinite-Dimensional Systems, Methods of

Francesco Dell'Isola, Fabio Di Cosmo

► **To cite this version:**

Francesco Dell'Isola, Fabio Di Cosmo. Lagrange Multipliers in Infinite-Dimensional Systems, Methods of. Encyclopedia of Continuum Mechanics, 2018, 9 p. hal-02099527

**HAL Id: hal-02099527**

**<https://hal.science/hal-02099527v1>**

Submitted on 15 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

## Lagrange Multipliers in Infinite-Dimensional Systems, Methods of

Francesco Dell'Isola<sup>1,2</sup> and  
Fabio Di Cosmo<sup>1</sup>

<sup>1</sup>International Research Center M&MoCS,  
University of L'Aquila, L'Aquila, Italy

<sup>2</sup>DISG, Università di Roma "La Sapienza",  
Roma, Italy

### Synonyms

Constrained mechanical systems

### Definitions

This entry will describe Lagrange multipliers method using a formulation which is valid for infinite-dimensional dynamical systems. The method of Lagrange multipliers is employed to deal with systems subject to constraints. The theoretical foundations of this method are presented, and a proof of the main theorem is illustrated for the relevant case of constraints defined on a Banach vector space.

### Introduction

Many physical systems change drastically their behavior when subject to kinematical constraints.

The mathematical models introduced for their description must therefore account for such effects. For instance, one can consider the example of a particle in a horizontal plane moving along a circular guideline: the presence of the guideline changes remarkably the effects of the forces applied on the particle and reduces its degrees of freedom from two to one.

An explicit description of the constraint can be obtained by introducing only one generalized Lagrangian coordinate, e.g., the angle with respect to the  $x$ -axis, and then equations of motion are derived by suitably particularizing the Hamilton-Rayleigh principle. On the other hand, one can effectively consider a superabundant set of Lagrangian coordinates and account for the effects of the constraints via the introduction of Lagrange multipliers in the formulation of the same principle. A modern version of the theory subjacent their use is shortly presented in this entry.

According to variational principles (see, e.g., Feynman et al. 1979; Lanczos 2012; Landau and Lifshitz 1982; Arnold 1978; dell'Isola and Placidi 2011), the evolution of a dynamical system is a stationary point of a certain functional, called action functional. In this framework, one can exploit some theorems coming from functional analysis and differential geometry to simplify the task of looking for the stationary points of an action functional in the presence of constraints. In this entry, the focus will be on Lagrange multipliers theorem, according to which the original stationarity problem is solved by

introducing a new functional which, in addition to the original one, contains the functions defining the set of constraints.

The roots of this method trace back to Lagrange who in his *Mécanique Analytique* (Lagrange 1811, 1815) presented many applications ranging from the description of the motion of particles constrained to given surfaces up to fluids in a pipe of a certain shape. The work by Lagrange was later continued by Piola who exploited systematically Lagrangian method to mechanical problems in his essays *Sull'applicazione de' principj della Meccanica Analitica del Lagrange ai principali problemi* (Piola 1825). Since then so many new applications have been introduced that it is useless to try to list the most important ones.

This entry will mainly focus on a modern formulation of Lagrange multipliers theorem using the language of differential geometry. However it is important to stress that the ideas behind these techniques are already fully contained in Lagrange's works. The mathematical discussion will be supported by simple applications, for finite-dimensional systems, in order to make the theoretical contents more understandable.

## Lagrange Multipliers Theorem

In this section, Lagrange multipliers theorem will be presented in a form which is suitable also for dynamical systems defined on infinite-dimensional configuration spaces, a situation which is frequently encountered in continuum mechanics. Before stating the theorem, it is worth recalling some notions which will be useful in the following. In particular, some basic elements of differential geometry of Banach manifolds will be introduced. In the exposition, the notation given in Abraham et al. (2001) will be adopted. Furthermore, it will be assumed that the reader is familiar with the general theory of Banach vector spaces (see Rudin 1991; Reed and Simon 1980).

However, for the sake of consistency, the starting point is the definition of a Banach manifold  $\mathcal{M}$ .

Let  $\mathcal{M}$  and  $E$  be a topological space and a Banach vector space, respectively. A chart on  $\mathcal{M}$  is a pair  $(U, \phi)$  where  $U$  is an open subset of  $\mathcal{M}$  and  $\phi: U \rightarrow \phi(U) \subset E$  is a bijection. A smooth atlas  $\mathcal{A}$  on  $\mathcal{M}$  is a family of charts  $\{(U_i, \phi_i), i \in I\}$  such

that:

- $\cup_{i \in I} U_i = \mathcal{M}$ .
- If two subsets  $U_i, U_j$  have a nonempty intersection, then the corresponding maps  $\phi_i, \phi_j$  must satisfy the following compatibility condition, i.e.,  $\phi_i \circ \phi_j^{-1} : \phi_j(U_i \cap U_j) \rightarrow \phi_i(U_i \cap U_j)$  is a smooth diffeomorphism in  $E$ .

Two atlases are equivalent if their union is still an atlas. An equivalence class of atlases  $\mathcal{D}$  is called a differentiable structure, and the union of all the atlases of the equivalence class is called maximal atlas.

The topological space  $\mathcal{M}$  equipped with the differentiable structure  $\mathcal{D}$  is called a **smooth Banach manifold**. If the functions  $\phi_i \circ \phi_j^{-1}$  are  $C^k$ -functions, the manifold will be called a  $C^k$ -differential manifold.

In other words, a smooth Banach manifold is a topological space which locally is smoothly equivalent to a Banach vector space, which may be either finite or infinite dimensional. In this entry, we will limit our considerations to manifold based on Banach spaces, but it is not the most general known possibility. Depending on the case of interest, one can model a neighborhood of a point in a manifold as a locally convex topological space (see, e.g., Michor 1980). Of interest is also the case in which  $E$  is actually a Hilbert spaces (see, e.g., Lang 1999).

Usually some further properties are assumed for the topological space  $\mathcal{M}$ , but there is no absolute agreement on these stipulations in the literature. For instance, it is common to assume that  $\mathcal{M}$  is second countable and Hausdorff.

A (super)-structure (in the technical sense of mathematical logic (see Hodges 2009)) which can be introduced after having defined a manifold is the tangent space to  $\mathcal{M}$  at a point  $x$ . Roughly speaking, a tangent vector is a possible velocity, in a possible motion of the system whose set of configurations is given by the manifold  $\mathcal{M}$ .

Let  $I \subset \mathbf{R}$  be a closed interval containing 0 in its interior. A differentiable curve is a  $C^1$ -map from  $I$  to the manifold  $\mathcal{M}$ . A map  $c: I \subset \mathbf{R} \rightarrow \mathcal{M}$  is a  $C^1$ -map at  $t_0$  if for any chart  $(U, \phi)$  with  $c(t_0) \in U$ , the function  $\phi \circ c: I \rightarrow \phi(U) \subset E$  has continuous derivative (indeed, if one changes the chart inside one atlas, the class of differentiability does not change. This circumstance actually motivates the definition of atlas).

An equivalence relation on this space of applications can be introduced. Two curves,  $c_1(t)$ ,  $c_2(t)$ , are tangent at the point  $x \in \mathcal{M}$  if, for any chart  $(U, \phi)$  with  $x \in U$ ,  $c_1(0) = c_2(0) = x$  and  $\left. \frac{d}{dt} (\phi \circ c_1) \right|_{t=0} = \left. \frac{d}{dt} (\phi \circ c_2) \right|_{t=0}$ . Obviously if this equality holds in one chart, it holds in all other charts of an atlas.

Elements of the tangent space  $T_x \mathcal{M}$  are the equivalence classes  $[c]_x$  of differentiable curves  $c(t)$  of  $\mathcal{M}$  which are tangent at  $x$ .

Informally speaking, the tangent space at a point is the space of the “generalized velocities” of the dynamical system which has  $\mathcal{M}$  as configuration space. It is worth remarking that one can choose as representative for any equivalence class the curve  $\tilde{c}$  such that  $(\tilde{c})(t) = \phi^{-1}(\phi(x) + t\mathbf{e})$ , with  $\mathbf{e} \in E$ . Consequently any tangent vector  $[c]_x$  at the point  $x$  is in one-to-one correspondence with the vector  $\mathbf{e} \in E$  satisfying

$$\mathbf{e} = \left. \frac{d}{dt} (\phi \circ \tilde{c}) \right|_{t=0}. \quad (1)$$

The union of the tangent spaces at every point of a manifold is a topological space which can be equipped with the structure of a Banach manifold, modeled on  $J := E \times E$ , and it is called the tangent bundle  $T\mathcal{M}$  of  $\mathcal{M}$ . Given a maximal atlas  $\mathcal{A}$  with charts  $(U_j, \phi_j)$  for  $\mathcal{M}$ , one can build an atlas  $\mathcal{I}\mathcal{A}$  for the tangent bundle  $T\mathcal{M}$ . Indeed, the open neighborhoods generating a topology in

$T\mathcal{M}$  are given by the union of sets  $A_x \subset T_x \mathcal{M}$ , which are the collections of the open sets  $A_x$  included in the tangent spaces at the points  $x \in U_j$ . Moreover it is easy to see that the maps  $\bar{\phi}_j: TU_j \rightarrow \phi(U) \times E \subset J$  defined as follows:

$$\bar{\phi}_j([c]_x) = (\phi_j(x), \mathbf{e}),$$

where  $\mathbf{e}$  has been defined in Eq. 1, actually form a set of charts in the atlas  $T\mathcal{A}$ .

Let  $(\mathcal{M}, \mathcal{D}_1)$  and  $(\mathcal{N}, \mathcal{D}_2)$  be two Banach manifolds based on the vector spaces  $E$  and  $F$ , respectively. A smooth map  $f: \mathcal{M} \rightarrow \mathcal{N}$  is a map such that for every  $x \in \mathcal{M}$  and admissible chart  $(V, \psi)$  of  $\mathcal{N}$  with  $f(x) \in V$ , for any admissible chart  $(U, \phi)$  of  $\mathcal{M}$  with  $x \in U$  and  $f(U) \subset V$ , the local representative

$$\tilde{f} = \psi \circ f \circ \phi^{-1}: \phi(U) \subset E \rightarrow \psi(V) \subset F$$

is smooth. The reader will understand here the reason for which the concept of manifolds and atlas has introduced. One has to be able to define the set of  $C^k$ -map from one manifold to another one. Actually it is sufficient to define differentiability in one chart in the domain of the function and in one chart of the codomain since the passage to all other charts can be obtained by composition with differentiable functions. If the local representative of a map is smooth with respect to a pair of charts, then it is smooth for every pair of charts in the corresponding maximal atlases. One can notice that the map  $\tilde{f}$  is a map between open subset of Banach spaces. Here and in the following, the superposed tilde will refer to the local representative of a given map between manifolds.

Given a  $C^1$ -map  $f: \mathcal{M} \rightarrow \mathcal{N}$ , the tangent map  $Tf: T\mathcal{M} \rightarrow T\mathcal{N}$  is a map from the tangent bundle of  $\mathcal{M}$  to the tangent bundle of  $\mathcal{N}$  which is defined as follows:

$$Tf([c]_x) = [f \circ c]_{f(x)},$$

i.e.,  $Tf$  maps the tangent vector  $[c]_x$  at  $x$  to the tangent vector  $[f \circ c]_{f(x)}$  at the point  $f(x)$ . The reader should notice that both members in the above definition are independent of the

particular curve in the equivalence class. In a pair of charts  $(TU, \bar{\phi})$ ,  $(TV, \bar{\psi})$  of  $T\mathcal{M}$  and  $T\mathcal{N}$ , respectively, with  $[c]_x \in TU$  and  $[f \circ c]_{f(x)} \in TV$ , the map  $Tf$  can be represented as follows:

$$T \tilde{f}(x, \mathbf{e}) = \left( \tilde{f}(x), D\tilde{f}(x) \cdot \mathbf{e} \right), \quad (2)$$

where  $\tilde{f}$  denotes the local representative of the function  $f$ .

Let us remark that for  $\mathcal{M}$ ,  $\mathcal{N}$  both finite-dimensional manifolds, with respective dimensions  $m$  and  $n$ , the local representative  $D\tilde{f}(x) \cdot \mathbf{e}$  denotes the action of the  $n \times m$  Jacobian matrix,  $D\tilde{f}(x)$ , on a tangent vector  $\mathbf{e} \in \mathbf{R}^m$ . A given  $C^1$ -map  $f : \mathcal{M} \rightarrow \mathcal{N}$  is called a submersion if the map  $T_x f$ , restriction of the map  $Tf$  in which  $x \in \mathcal{M}$  is kept fixed, is surjective with split kernel for all  $x \in \mathcal{M}$ . A map has split kernel if the kernel  $K \subset T_x \mathcal{M}$  is a closed subspace which is complemented by a closed subset  $H$ , i.e.,  $T_x \mathcal{M} = K \oplus H$ . This property is automatically satisfied if  $E$  is either a finite-dimensional vector space or a Hilbert vector space.

After this very short summary of differential geometry, it is possible to state Lagrange multipliers theorem as follows.

### Theorem 1 Lagrange Multipliers Theorem

Let  $F, F^*$  be a Banach vector space and its dual space, respectively,  $\mathcal{M}$  a smooth manifold,  $g : \mathcal{M} \rightarrow F$  a smooth submersion,  $\mathcal{N} = g^{-1}(0)$ , and  $f : \mathcal{M} \rightarrow \mathbf{R}$  be a  $C^r$ -function with  $r \geq 1$ . The point  $n \in \mathcal{N}$  is a critical point for the restriction  $f|_{\mathcal{N}}$  if there exists a  $\lambda \in F^*$ , called a Lagrange multiplier, such that  $n$  is a critical point of  $f - (\lambda \circ g)$ .

It is worth stressing once more that this form of the theorem does not require finite-dimensional manifolds or vector spaces. However in the next section, a proof of the theorem will be illustrated which is valid in the finite-dimensional case where the geometrical content can be more easily visualized. In the entry “► [Lagrange Multipliers in Infinite Dimensional Systems, Examples of Applications](#),” such a general setting is needed in the theory of generalized continua.

With an abuse of notation, in what follows a manifold will be denoted only with the symbol  $\mathcal{M}$ , without explicitly referring to its differentiable structure  $\mathcal{D}$ , which is assumed chosen once forever.

## Finite-Dimensional Case

The goal of this section consists in proving the previous theorem when  $\mathcal{M}$  is a finite-dimensional manifold, with dimension  $m$ , and  $F$  is a finite-dimensional vector space, with dimension  $k < m$ , i.e.,  $F = \mathbf{R}^k$ .

**Proof** Let  $f : \mathcal{M} \rightarrow \mathbf{R}$  be a function of class  $C^r$  ( $\mathcal{M}$ ), with  $r \geq 1$ , and  $g : \mathcal{M} \rightarrow \mathbf{R}^k$  be a smooth submersion. Obviously  $g$  is a smooth submersion in a point of its domain if the rank of its Jacobian matrix is maximal.

A point  $x \in \mathcal{M}$  is said a critical point of  $f$  if  $T_x f = 0$ , which in a chart, can be written as  $Df(x) = 0$ . If one considers the restriction  $f|_{\mathcal{N}}$  of the function  $f$  to the submanifold  $\mathcal{N} = g^{-1}(0)$ , the point  $n \in \mathcal{N}$  is a critical point of this restriction if  $T_n f$  vanishes on any vector tangent to the submanifold  $\mathcal{N}$ . In a local chart, this condition means that  $Df(x) \cdot \mathbf{v} = 0$  for all  $\mathbf{v} \in T_n \mathcal{N}$ . It is clear here that one can limit himself to assume that for every  $n \in \mathcal{N}$  the rank of the Jacobian matrix of  $g$  is maximal.

Therefore, the first step to prove Lagrange multiplier theorem is to characterize the tangent space to the submanifold  $\mathcal{N} = g^{-1}(0)$ . Let  $n$  be a point of  $\mathcal{N}$  and let  $(U, \phi)$  be a chart of  $\mathcal{M}$  with  $n \in U$ . A local expression for the submersion  $g : \mathcal{M} \rightarrow \mathbf{R}^k$  is the following:

$$\begin{aligned} y^\mu &= \tilde{g}^\mu(x), \quad \mu = 1, \dots, k \quad \text{and} \\ x &= (x^1, \dots, x^m). \end{aligned} \quad (3)$$

By definition, the points which belong to the submanifold  $\mathcal{N}$  have coordinates  $x^j$  which satisfy the equations  $g^\mu(x) = 0$ . Therefore a curve  $c : I \subset \mathbf{R} \rightarrow \mathcal{N}$ , which can be represented as the  $n$ -ple of functions  $x(t) = (x^1(t), \dots, x^m(t))$ , fulfills the condition  $\tilde{g}^\mu(x(t)) = 0$ , and consequently we have that:

$$\left. \frac{d}{dt} (\tilde{g}^\mu(x(t))) \right|_{t=0} = \left. \frac{\partial \tilde{g}^\mu}{\partial x^j} \right|_{x(0)} \left. \frac{dx^j}{dt} \right|_{t=0} = 0. \quad (4)$$

This means that tangent vectors to the submanifold  $\mathcal{N} = g^{-1}(0)$  are in the kernel of the linear application  $T_n g$ , for all  $n \in \mathcal{N}$ .

It is now possible to prove Lagrange multipliers theorem. Indeed let  $n$  be a critical point for  $f|_{\mathcal{N}}$ . Then if one considers the local expression  $\tilde{f}(x(t))$  of the restriction of the function  $f$  to a curve  $c$  on  $\mathcal{N}$  passing through  $n$  when  $t = 0$ , one gets that

$$\left. \frac{d}{dt} \tilde{f}(x(t)) \right|_{t=0} = \left. \frac{\partial \tilde{f}}{\partial x^j} \right|_{x(0)} \left. \frac{dx^j}{dt} \right|_{t=0} = 0$$

for all vectors which are tangent to the submanifold  $\mathcal{N}$ . Since these vectors satisfy Eq. 4 and  $g$  is a submersion (which implies that the Jacobian  $\frac{\partial \tilde{g}^\mu}{\partial x^j}$  has maximal rank), one can conclude that there exist  $(\lambda_1, \dots, \lambda_k) \in (\mathbf{R}^k)^* \equiv \mathbf{R}^k$  such that:

$$\left. \frac{\partial \tilde{f}}{\partial x^j} \right|_n = \lambda_\mu \left. \frac{\partial \tilde{g}^\mu}{\partial x^j} \right|_n.$$

Therefore  $n$  is a critical point for the function  $f - \lambda \circ g$ .

Vice versa if there exists a  $\lambda \in (\mathbf{R}^k)^*$  such that  $n$  is a critical point of  $f - \lambda \circ g$ , it means that locally  $D\tilde{f}(n) = \lambda \circ D\tilde{g}(n)$ . Since all the vectors which are tangent to  $\mathcal{N}$  are in the kernel of  $D\tilde{g}(n)$ , also  $D\tilde{f}(n)$  vanishes on these vectors. Therefore  $n$  is a critical point for the function  $f|_{\mathcal{N}}$ .

**Example 1** To clarify the concept, it is worth starting with a very simple example: finding the critical points of a function  $f : \mathbf{R}^3 \rightarrow \mathbf{R}$ , restricted to a given line. The function is  $f(x, y, z) = xy + yz$  with the two constraints which are  $g_1 = x + 2y - 6 = 0$  and  $g_2 = x - 3z = 0$ .

According to Lagrange multipliers theorem, one has to find the critical points of the function  $f - \lambda_1 g_1 - \lambda_2 g_2 = zy + yz - \lambda_1(x + 2y - 6) - \lambda_2(x - 3z)$ . Therefore the equations to solve are:

$$\begin{aligned} y - \lambda_1 - \lambda_2 &= 0 \\ x + z - 2\lambda_1 &= 0 \\ y + 3\lambda_2 &= 0 \end{aligned}$$

subjected to the constraints  $g_1 = 0$  and  $g_2 = 0$ .

A straightforward computation shows that the only critical point  $P$  has coordinates  $P \equiv (3, \frac{3}{2}, 1)$ .

## Constraints in Banach Spaces

The advantage deriving from the application of Lagrange multipliers theorem was evident already to Gabrio Piola (1825), who strongly supported the use of this method in mechanics.

This advantage is also evident when the manifold  $\mathcal{M}$  reduces to a Banach or Hilbert vector space, as the presence of a linear structure allows for the exploitation of all the well-established techniques of Linear Operator Theory. Consequently, many applications of Lagrange multipliers theorem concern dynamical systems subject to this kind of constraints.

In a Banach vector space, there exists a globally defined chart. Therefore the expression of the tangent map  $T_x f$  at a point  $x \in \mathcal{M}$  can be replaced by its local expression  $Df(x)$ , in terms of the differential of the function.

In this section, therefore, the general results of Theorem 1 will be adapted to the case of a real-valued functional  $f$  defined on a Banach space  $X$  and subject to a constraint  $g: X \rightarrow Z$ , where  $Z$  is another Banach space. Both,  $X$  and  $Z$ , can be infinite-dimensional vector spaces. In this case, some proofs will be shown, which make use of fundamental theorems of linear analysis, like the implicit function theorem. More details about the topic can be found in the book (Zeidler 1984a).

### Submersions

This paragraph aims at showing that when  $G: X \rightarrow Z$  is a  $C^1$ -submersion (see section Lagrange Multipliers Theorem), the set of points  $n \in X$  such that  $G(n) = 0$  form a closed submanifold of the Banach space  $X$ . Since the proof of this fact is based on the implicit function theorem, we start with the statement of this useful result of analysis.

## Theorem 2 Implicit Function Theorem

Let  $X, Y, Z$  be Banach spaces. Suppose that:

- $G: U(x_0, y_0) \subset X \times Y \rightarrow Z$  is a map defined on an open subset  $U(x_0, y_0)$  and  $G(x_0, y_0) = 0$ .
- The partial Fréchet derivative  $D_y G$  exists on  $U(x_0, y_0)$  and is bijective.
- $G$  and  $D_y G$  are continuous at  $(x_0, y_0) \in X \times Y$ .

Then:

- There exist two constants  $r, r_0$  such that, for every  $x \in X$  satisfying  $\|x - x_0\|_X \leq r_0$ , there is only one  $y(x) \in Y$  for which  $\|y(x) - y_0\|_Y \leq r$  and  $G(x, y(x)) = 0$ ;
- if  $G$  is a  $C^m$ -map on a neighborhood of  $(x_0, y_0)$ , then  $y(\cdot)$  is also a  $C^m$ -map on a neighborhood of  $x_0$ .

The function  $y(\cdot)$  is called implicit function. For a proof of this theorem, see Zeidler (1984b).

As mentioned above, before considering Lagrange multipliers theorem, a preliminary results, which require  $G$  to be a submersion, will be proved:

## Theorem 3

Let  $X, Z$  be two Banach spaces. If  $G: X \rightarrow Z$  is a  $C^1$ -submersion at  $u_0 \in X$ , Fréchet differentiable in a neighborhood of  $u_0$ , with  $G(u_0) = 0$ , and  $\mathcal{M}$  denotes the set of points  $\{n \in X: G(n) = 0\}$ , then:

- There exists a  $C^1$ -diffeomorphism  $\phi: V \subset K := \text{Kern}(DG(u_0)) \rightarrow O \subset \mathcal{M}$ , with  $V$  open set including 0, onto an open neighborhood  $O$  of  $u_0 \in \mathcal{M}$ . Moreover,  $\phi(h) = u_0 + h + o(\|h\|)$  as  $h \rightarrow 0$ .
- If  $G$  is a  $C^1$ -submersion for all  $n \in X$  such that  $G(n) = 0$ , then  $\mathcal{M}$  is a  $C^1$ -manifold. If, in addition,  $G$  is a  $C^r$ -map, then  $\mathcal{M}$  is a  $C^r$ -manifold.
- The tangent space  $T_{u_0} \mathcal{M}$  at the point  $u_0 \in \mathcal{M}$  is the kernel  $K$  of the differential  $DG(u_0)$ .

**Proof** Since  $G$  is a submersion, its kernel splits. Therefore  $X = K \oplus Y$ , and there exists a projec-

tion operator  $P: X \rightarrow K$  such that any element  $u \in X$  can be uniquely decomposed into the sum  $u = Pu + (\mathbf{I} - P)u = u_1 + u_2$ .

Let  $W(u_1, u_2) = G(u_0 + u_1 + u_2)$  be an auxiliary function. In order to prove the first point in the theses, one can apply the implicit function theorem to this function. Therefore one has to control if the hypotheses of the implicit function theorem are satisfied:

- $W: U(0, 0) \subset K \times Y \rightarrow Z$  satisfies  $W(0, 0) = G(u_0) = 0$ , where  $U(0, 0)$  is a neighborhood of the origin  $(0, 0)$ .
- $D_{u_1} W(u_1, u_2)k = DG(u_0 + u_1 + u_2)k$ ,  $\forall k \in K$ , and  $D_{u_2} W(u_1, u_2)h = DG(u_0 + u_1 + u_2)h$ ,  $\forall h \in Y$ , are continuous in  $U(0, 0)$ . Thus,  $W$  is Fréchet differentiable on  $U(0, 0)$ .
- The operator  $A = D_{u_2} W(0, 0)$  is linear, continuous, and bijective. Indeed, since  $A$  coincides with the restriction of  $DG(u_0)$  to  $Y$ ,

$$DG(u_0)h = 0, \quad h \in Y \Rightarrow h \in K \Rightarrow h = 0,$$

and the range  $R(A)$  is equal to  $Z$ .

According to the implicit function theorem, there exist two numbers  $r_0, r$  and a neighborhood  $V_0(0, 0) \subset K \times Y$  such that for each  $u_1 \in K$  satisfying  $\|u_1\| \leq r_0$ , there is only one  $\psi(u_1) \in Y$ , obeying  $\|\psi(u_1)\| \leq r$ , with  $(u_1, \psi(u_1)) \in V_0(0, 0)$  and

$$W(u_1, \psi(u_1)) = 0. \quad (5)$$

Furthermore  $\psi$  is continuously Fréchet differentiable. Therefore the derivative of Expression 5 satisfies:

$$D_{u_1} W(0, 0) + D_{u_2} W(0, 0) D\psi(0) = 0. \quad (6)$$

Since  $D_{u_1} W(0, 0) = 0$ , then  $D_{u_2} W(0, 0) D\psi(0) = 0$  and consequently  $D\psi(0) = 0$ . As  $\psi(0) = D\psi(0) = 0$ , one has that  $\psi(u_1) = o(\|u_1\|)$  as  $u_1 \rightarrow 0$ .

Let  $\phi$  be the map  $\phi(u_1) = u_0 + u_1 + \psi(u_1)$ . This is a homeomorphism of an open neighborhood  $V$  of  $0 \in K$  onto an open neighborhood

$O$  of  $u_0 \in \mathcal{M}$ . The inverse map is obtained by means of the projection operator  $P$ . Indeed, for a sufficiently small  $\varepsilon > 0$ , one has:

$$\begin{aligned} \|u - u_0\| < \varepsilon, \quad G(u) = 0 &\Rightarrow u = u_0 + u_1 + u_2, \\ u_1 = P(u - u_0) = \phi^{-1}(u) \\ \|u_1\| < r_0 &\Rightarrow u_2 = \psi(u_1) \Rightarrow u = \phi(u_1). \end{aligned} \quad (7)$$

If  $G$  is a  $C^r$ -map in an open neighborhood of  $u_0$ , then  $W$  is a  $C^r$ -map and consequently the maps  $\psi$  and  $\phi$ .

In order to prove the  $C^1$ -manifold structure, let  $\bar{u}_0$  be a second point such that  $G(\bar{u}_0) = 0$ , and let  $\bar{\phi}$  be the corresponding homeomorphism defined above. Suppose that the range,  $O$  and  $\bar{O}$ , of  $\phi$  and  $\bar{\phi}$  has a nonempty intersection. If  $u \in O \cap \bar{O}$  has the following two representations:

$$\begin{aligned} u = \phi(h), \quad h \in K, \quad \text{and} \quad u = \bar{\phi}(\bar{h}) \\ \bar{h} \in K, \end{aligned} \quad (8)$$

then the relation

$$h = P(u - u_0) = P(\bar{\phi}(\bar{h}) - u_0) \quad (9)$$

follows. Since  $P \circ \bar{\phi}$  is continuously Fréchet differentiable,  $\mathcal{M}$  is a  $C^1$ -manifold, modeled on the Banach space  $K$  and having charts  $(O_j, \phi_j^{-1})$ .

If  $G$  is a  $C^r$ -map, then  $\bar{\phi}$  and  $P \circ \bar{\phi}$  are  $C^r$ -maps and  $\mathcal{M}$  is a  $C^r$ -manifold.

Finally, let  $[u]_{u_0}$  be a vector tangent to the space  $\mathcal{M}$  and  $u(t)$  a curve in its equivalence class. If  $(O, \phi^{-1})$  is a local chart of  $\mathcal{M}$ , then  $G(\tilde{u}(t)) = 0$  implies that  $DG(u_0)h = 0$ , where  $h = \left. \frac{d}{dt} (\phi^{-1} \circ u) \right|_{t=0}$ .

Conversely, if  $DG(u_0)h = 0$  holds, then  $u(t) = \phi(th)$  is a representative of the equivalence class associated with the tangent vector  $[u]_{u_0}$ , according to the previous definition of the homeomorphism  $\phi$ .

### Lagrange Multipliers Theorem in Banach Spaces

The main result obtained in the previous paragraph is that the inverse image of a regular value

of a submersion is a closed submanifold of a Banach space. A point  $n$  in a manifold  $\mathcal{N}$  is called a regular value for the map  $G: \mathcal{M} \rightarrow \mathcal{N}$  if for each point  $m \in G^{-1}(n)$ ,  $T_m G$  is surjective with split kernel.

The next step is proving the Lagrange multiplier theorem for a real-valued function on a Banach space with a constraint given by a  $C^1$ -submersion.

Differently from the statement of the theorem exposed in Section 2, in this case, it is possible to provide also a sufficient condition for the existence of extrema of a function  $f$  on a constrained submanifold in a Banach space.

However, before proving the main theorem of this section, an additional result is needed:

#### Theorem 4

Let  $X, Y$  be Banach spaces over the field  $\mathbf{K}$ , with  $\mathbf{K} = \mathbf{R}$  or  $\mathbf{C}$ . Let  $B: X \rightarrow \mathbf{K}$  and  $A: X \rightarrow Y$  be two continuous linear operators, with  $R(A) = Y$ , where  $R(A)$  is the range of the operator  $A$ .

If  $Bh = 0 \forall h \in X$  such that  $Ah = 0$ , then there exists a unique  $\Lambda \in Y^*$  such that:

$$Bk - \Lambda(Ak) = 0, \quad \forall k \in X. \quad (10)$$

**Proof** According to the closed range theorem,  $R(A^*) = N(A)^\perp$ . By hypothesis  $B \in N(A)^\perp$ , and thus there exists a  $\Lambda \in Y^*$  such that  $B = A^* \Lambda$ . Consequently one has:

$$Bk = (A^* \Lambda)(k) = \Lambda(Ak), \quad \forall k \in X. \quad (11)$$

Since  $R(A) = Y$ , then  $N(A^*) = R(A)^\perp = \{0\}$ , i.e.,  $A^*$  is injective and  $\Lambda$  is uniquely determined by  $B$ .

Now the following version of Lagrange multipliers theorem can be proved:

#### Theorem 5

Let  $f: X \rightarrow \mathbf{R}$  be a Fréchet differentiable function at  $u_0$  and  $G: X \rightarrow Y$  a  $C^1$ -submersion at  $u_0$ . Let  $k^r$  denote the  $r$ -ple  $k^r = \underbrace{(k, \dots, k)}_{r\text{-times}} \in$

$$\underbrace{X \times \dots \times X}_{r\text{-times}}.$$



Then:

- If  $f$  restricted to  $G^{-1}(0) =: \mathcal{M}$  has a local minimum at  $u_0$ , then there exists a  $\Lambda \in Y^*$  such that

$$Df(u_0)k = \Lambda(DG(u_0)k) \quad \forall k \in X. \quad (12)$$

- $f$  has a strict local minimum at  $u_0$  when:

1.  $f$  and  $G$  are  $n$ -times continuously Fréchet differentiable in an open neighborhood of  $u_0$ , with  $n \geq 2$ .
2. There exist a number  $c > 0$  and a functional  $\Lambda \in Y^*$  such that

$$Df^{(r)}(u_0) \cdot k^r - \Lambda(DG^{(r)}(u_0) \cdot k^r) = 0 \quad (13)$$

$$r = 1, \dots, n = 1$$

$$Df^{(n)}(u_0) \cdot h^n - \Lambda(DG^{(n)}(u_0) h^n) \geq c \|h\|^n. \quad (14)$$

$\forall k \in X$  and  $\forall h$  satisfying  $DG(u_0)h = 0$ .

**Proof** Consider a tangent vector  $[c]_{u_0}$  at the point  $u_0 \in \mathcal{M}$  and the curve  $\tilde{c}(t) = u_0 + ht$ , with  $h \in K$ , as representative of its equivalence class. Since  $G(\tilde{c}(t)) = 0$ ,  $DG(u_0)h = Ah = 0$ . On the other hand, since  $u_0$  is a minimum of the functional  $f$ , the function  $F(t) = f(\tilde{c}(t))$  has a local minimum at  $t = 0$ . Therefore  $\frac{dF}{dt}(0) = Df(u_0)h = Bh = 0$  for all  $h \in K$ . By Theorem 4, one gets the existence of the Lagrange multiplier  $\Lambda \in Y^*$ .

In order to prove the sufficient condition, one can consider the case of  $f$  twice differentiable. Define

$$H(u) = f(u) - \Lambda(G(u)). \quad (15)$$

For  $r = 1, 2$ , we get

$$DH^{(r)}(u_0) \cdot k^r = Df^{(r)}(u_0) \cdot k^r - \Lambda(DG^{(r)}(u_0) \cdot k^r).$$

Taylor expansion for the function  $H(u)$  can be written as follows:

$$H(u) - H(u_0) = DH^{(2)}(u_0)(u - u_0)^2 + o(\|u - u_0\|_X^2). \quad (16)$$

Choose  $u = \phi(h)$  for  $h$  in a small neighborhood of zero,  $V \subset K$ . Then  $G(u) = 0$  and  $H(u) = f(u)$ . Since  $\phi(h) = u_0 + h + o(\|h\|_X)$ , one has that for all  $h \in V$

$$f(\phi(h)) - f(u_0) = DH^{(2)}(u_0)h^2 + o(\|h\|_X^2) \geq c\|h\|_X^2 + o(\|h\|_X^2), \quad (17)$$

as  $h \rightarrow 0$ .

Consequently  $f$  has a local minimum at  $u_0$  in  $\mathcal{M}$ .

### Example

As a very simple example, one can consider a thermodynamical system with a finite number  $N$  of possible states. Let  $\Sigma_1, \Sigma_2, \dots, \Sigma_N$  denote the possible states of the system and  $E_1, E_2, \dots, E_N$  the corresponding energy values. Finally  $p_i$  will be the probability of finding the system in the state  $\Sigma_i$ .

If  $K = \{p \in \mathbf{R}^N: 0 \leq p_i \leq 1, i = 1, 2, \dots, N\}$ , the equilibrium state of the system can be found as the maximum over  $K$  of the entropy function

$$S(\vec{p}) = -k_B \sum_{i=1}^N p_i \ln(p_i) \quad (18)$$

subject to the following constraints:

$$\sum_{i=1}^N p_i = 1 \quad (19)$$

$$\sum_{i=1}^N p_i E_i = E \quad (20)$$

for fixed  $E$ . These constraints represent the fact that the vector  $\vec{p}$  must define a probability density and that the system under analysis has a fixed average energy. This can happen if, for instance, one is considering a system in contact with a thermal reservoir with a very weak interaction.

One can easily solve this optimization problem with constraints by applying the previously discussed Lagrange multiplier theorem. Therefore one has to find the stationary points of the new function:

$$L(\vec{p}, \Lambda_1, \Lambda_2) = S(\vec{p}) + \Lambda_1 \left( \sum_{i=1}^N p_i - 1 \right) + \Lambda_2 \left( \sum_{i=1}^N p_i E_i - E \right). \quad (21)$$

Stationary points satisfy the Euler-Lagrange equations:

$$\frac{\partial L}{\partial p_i} = -k_B (\ln(p_i) + 1) = \Lambda_1 + \Lambda_2 E_i$$

and a solution of these equations can be written as

$$p_i = c_1 e^{c_2 E_i} \quad (22)$$

with  $c_1$  and  $c_2$  constants to be chosen such that the two constraints 19 and 20 are satisfied. In statistical mechanics, one sets the value  $c_2 = \frac{1}{k_B T}$  where  $T$  is interpreted as the absolute temperature of the thermodynamical system.

It is also very easy to show that this stationary point is a maximum, since the Hessian matrix can be written as

$$H(\vec{p}) = \frac{\partial^2 L}{\partial p_i \partial p_j} = -k_B \frac{1}{p_i} \delta_{ij},$$

which is negative definite.

## References

- Abraham R, Marsden JE, Ratiu T (2001) *Manifolds, tensor analysis, and applications*. Springer, New York
- Arnold VI (1978) *Mathematical methods of classical mechanics*. Springer, Berlin
- dell'Isola F, Placidi L (2011) Variational principles are a powerful tool also for formulating field theories. In: dell'Isola F, Gavriluk SL (eds) *Variational models and methods in solid and fluid mechanics*. Springer, Wien
- Feynman RP, Leighton RB, Sands M (1979) *The Feynman lectures on physics, vol. 2: Mainly electromagnetism and matter*. Addison-Wesley, Reading
- Hodges W (2009) Functional modelling and mathematical models. In: Meijers A (ed) *Philosophy of technology and engineering sciences. Handbook of the philosophy and science, 9*. Elsevier, Oxford
- Lagrange JL (1811) *Mécanique Analytique vol. 1*. M.me V. Courcier, Paris
- Lagrange JL (1815) *Mécanique Analytique vol. 2*. M.me V. Courcier, Paris
- Lanczos C (2012) *The variational principles of mechanics*. Courier Corporation, New York
- Landau LD, Lifshitz EM (1982) *Course of theoretical physics, vol. 1: Mechanics*. Butterworth-Heinemann, Oxford
- Lang S (1999) *Fundamentals of differential geometry*. Springer, New York
- Michor PW (1980) *Manifolds of differentiable mappings*. Shiva Publishing Limited, Orpington
- Piola G (1825) *Sull'applicazione de' principj della meccanica analitica del Lagrange ai principali problemi*. Dall'Imp/Reggia Stamperia, Milano
- Reed M, Simon B (1980) *Methods of modern mathematical physics, vol.1: Functional analysis*. Academic, San Diego
- Rudin W (1991) *Functional analysis*. McGraw-Hill, Singapore
- Zeidler E (1984a) *Nonlinear functional analysis and its applications III: variational methods and optimization*. Springer, New York
- Zeidler E (1984b) *Nonlinear functional analysis and its applications I: fixed-point theorems*. Springer, New York