THE EULER-LAGRANGE EQUATION AND ITS INVARIANCE PROPERTIES

Héctor J. Sussmann

November 8, 2000

1. THE EULER-LAGRANGE EQUATION

1.1. Standard calculus of variations problems. The calculus of variations deals mainly with optimization problems involving a function

$$(x, \dot{x}) \mapsto L(x, \dot{x}, t)$$

of "position, velocity, and time." Nowadays, this function is usually called the *Lagrangian.* In contemporary mathematics, the "position"—or "configuration," or "state"[1]—x is usually taken to be a vector[2], $x = (x^1, x^2, \ldots, x^n)$ in n-dimensional Euclidean space $\mathbb{R}^n$, and the velocity $\dot{x} = (\dot{x}^1, \dot{x}^2, \ldots, \dot{x}^n)$ is then also a vector in $\mathbb{R}^n$.

REMARK 1.1.1. In the earliest work, n was usually one or two or three. Considering higher dimensions is natural in today's mathematics, but the idea that one might as well write $(x^1, x^2, \ldots, x^n)$ rather than x, or (x, y), or (x, y, z), and "do everything in n-dimensions, with n completely arbitrary," was by no means natural to the founders of the subject and took a long time to evolve. Lagrange, for example, devotes a lot of space in his *Mécanique Analytique* to persuading the reader that a system of N point particles in 3-dimensional space can be thought of as a single point evolving in a space of $3N$ coordinates. $\diamondsuit$

REMARK 1.1.2. One can also consider problems in which x is a vector in an infinite-dimensional space, and a lot of the current activity deals with such questions. However, *here we will only discuss finite-dimensional problems.* $\diamondsuit$

The objective of a typical calculus of variations problem is to find a curve $t \mapsto \xi(t)$ in x-space that minimizes the integral

$$I(\xi) = \int_a^b L(\xi(t), \dot{\xi}(t), t)dt$$

of the Lagrangian, among all curves[3]

---

[1]We will point out below that one should really distinguish between the *configuration q* of a physical system and a *coordinate representation* in terms of an n-tuple $x = (x^1, x^2, \ldots, x^n)$ of numbers, but at this point we will ignore the distinction

[2]We follow the modern convention of using superscripts for "contravariant indices," i.e., components of vectors, and subscripts for "covariant indices," i.e., components of covectors.

[3]In twentieth century mathematics, "all curves" means, of course, all curves in some suitably chosen function space, such as that of all absolutely continuous curves, or that

For example, one could specify a time interval $[a, b]$, an initial point $\bar{x}$, and a terminal point $\hat{x}$, and seek to minimize I in the class of all curves $\xi : [a, b] \to \mathbb{R}^n$ sucht that $\xi(a) = \bar{x}$ and $\xi(b) = \hat{x}$. This is a "standard calculus of variations problem":

$$
\boxed{
\begin{array}{rl}
\multicolumn{2}{c}{\textbf{THE} \quad \textbf{PROBLEM} \quad \mathcal{P}(L, a, b, \bar{x}, \hat{x})} \\[2ex]
\text{Given} & L,\ a,\ b,\ \bar{x},\ \hat{x}\,, \\[2ex]
\text{minimize} & I = \int_a^b L(\xi(t), \dot{\xi}(t), t)dt\,, \\[2ex]
\text{subject to} & \xi(a) = \bar{x} \ \text{ and } \xi(b) = \hat{x}\,.
\end{array}
}
\tag{1.1}
$$

Alternatively, one may think that the use of the symbol $\dot{x}$ as a variable is a little bit confusing, and prefer to avoid using it. In order to make it transparently clear that L is "a function of three independent variables"—namely, the configuration point x, the velocity vector u, and the time t—one would use a letter such as u rather than $\dot{x}$ as the name for the velocity variable. This leads to regarding the Lagrangian as a function

$$\Omega \times \mathbb{R}^n \times [a, b] \ni (x, u, t) \mapsto L(x, u, t) \tag{1.2}$$

and rewriting (1.1) in the equivalent form

$$
\boxed{
\begin{array}{rl}
\multicolumn{2}{c}{\textbf{THE} \quad \textbf{PROBLEM} \quad \mathcal{P}(L, a, b, \bar{x}, \hat{x})} \\[2ex]
\text{Given} & L,\ a,\ b,\ \bar{x},\ \hat{x}\,, \\[2ex]
\text{minimize} & I = \int_a^b L(\xi(t), \eta(t), t)dt\,, \\[2ex]
\text{subject to} & \xi(a) = \bar{x}\,,\ \xi(b) = \hat{x}\,, \\[2ex]
\text{and} & \dot{\xi}(t) = \eta(t) \ \text{for} \ a \le t \le b\,.
\end{array}
}
\tag{1.3}
$$

The vector variable x is the *state*, or *configuration*, of the problem (1.1) or (1.3). The set Ω of all possible values of x is the *configuration space*, or *state space*, of the problem. For the time being, we will require Ω to be an *open subset of a Euclidean space* $\mathbb{R}^n$. Later on, we will allow Ω to be a more general *differentiable manifold*, for example a spherical surface.

So we will stipulate that

- the configuration space Ω is an open subset of $\mathbb{R}^n$;
- the variable u—the "velocity"—takes values in $\mathbb{R}^n$;
- L is a real-valued function on $\Omega \times \mathbb{R}^n \times [a, b]$.

---

of all Lipschitz curves. The choice of function space can sometimes be crucial, as will be seen when we discuss the "Lavrentiev phenomenon" later on.

REMARK 1.1.3. A minimization problem of the form (1.1)—or (1.3)—is a *point-to-point* calculus of variations problem, because the endpoint constraint on the trajectory $\xi(\cdot)$ specifies that $\xi(\cdot)$ has to start at a given point $\bar{x}$ and end at another given point $\hat{x}$. One can of course consider, both for the calculus of variations and for optimal control, more general *set-to-set* problems, in which the endpoint constraints are

$$\xi(a) \in S_1 \quad \text{and} \qquad \xi(b) \in S_2\,, \tag{1.4}$$

where S_1, S_2 are two given subsets of Ω. Even more generally, one can consider "mixed" endpoint constraints of the form

$$(\xi(a), \xi(b)) \in S\,, \tag{1.5}$$

where S is a given subset of $\Omega \times \Omega$. (For example, the constraint that ξ be a closed curve—i. e., that $\xi(b) = \xi(a)$—would be expressed by taking $S = \{(x, x) : x \in \Omega\}$.) Such more general problems have been studied since the very beginning of the calculus of variations, and give rise to interesting questions about *transversality conditions*. At this point, *we will only deal with calculus of variations (and optimal control) point-to-point problems*. This means that, for the time being, we are leaving out the whole area of transversality conditions. $\diamondsuit$

1.2. The Euler-Lagrange equation. We now look for *necessary conditions* for a given curve $t \mapsto \xi_*(t)$ to be a solution of the problem (1.1).

So we will be dealing with a problem $\mathcal{P}(L, a, b, \bar{x}, \hat{x})$ as before, together with a given curve ξ_*. The curve ξ_* will be referred to, in this kind of discussion, as "the reference curve," meaning "the curve that we are examining in order to decide if it is a solution of the minimization problem."

The general condition known today as the "Euler-Lagrange equation" was derived by *Leonhard Euler*[4] (1707-1793) and *Joseph-Louis Lagrange*[5] (1736-1813).

---

[4]Euler entered the University of Basel in 1720 at the age of 13. There he studied the methods developed by the Bernoullis for the study of isoperimetric problems, but soon became convinced that there was a need for a general theory, that would go beyond the solutions of specific problems. He began to develop his approach in 1731, and by 1740 he had achieved his goal and come up with a general theory. He published his results in 1744, in a book entitled *Methodus inveniendi maximi minime propriate gaudentes sive solutio problematis isoperimetrici latissimo sensu accepti* (A method for discovering curved lines having a maximum or minimum property or the solution of the isoperimetric problem taken in its widest sense). There he gave a general procedure for writing down what became known as *Euler's equation*. The words "calculus of variations" do not yet appear in the *Methodus inveniendi*. They were coined by Euler later, in 1760, as a name for Lagrange's method that Euler had decided to adopt in lieu of his own discretization approach, and appeared in print in 1766.

[5]Lagrange was a 19-year-old living in Turin when, on 12 August 1755, he wrote to Euler a brief letter to which was attached an appendix containing mathematical details of a new idea. He introduced a new method that could turn out the necessary condition of Euler, and more, almost automatically. After seeing Lagrange's work, Euler dropped his own method, espoused that of Lagrange, and renamed the subject the *calculus of variations*. In the summary to his first paper using variations, Euler says "Even though the author of this had meditated a long time and had revealed to friends his desire yet the glory of first discovery was reserved to the very penetrating geometer of Turin LA GRANGE, who having used analysis alone, has clearly attained the same solution which the author had deduced by geometrical considerations."

The Euler-Lagrange equation says[6] that the identity

$$\boxed{\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = \frac{\partial L}{\partial x}} \tag{1.6}$$

must be satisfied along the curve ξ_*.

Equation (1.6) makes perfect sense and is a necessary condition for optimality for a vector-valued variable x as well as for a scalar one. It can be written as a system:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\right) = \frac{\partial L}{\partial x^i}\,, \qquad i = 1, \ldots, n\,. \tag{1.7}$$

Alternatively, we can regard Equation (1.6) as a vector identity, in which

$$x = (x^1, \ldots, x^n)$$

denotes an n-dimensional vector, and the expressions

$$\frac{\partial L}{\partial x}\,, \qquad \frac{\partial L}{\partial \dot{x}}$$

stand for the n-tuples

$$\left(\frac{\partial L}{\partial x^1}, \ldots, \frac{\partial L}{\partial x^n}\right), \qquad \left(\frac{\partial L}{\partial \dot{x}^1}, \ldots, \frac{\partial L}{\partial \dot{x}^n}\right).$$

A modern mathematician might be troubled by the use of $\dot{x}$ both as an "independent variable" and as a function of time evaluated along a trajectory, and might prefer to write (1.6) as

$$\boxed{\frac{d}{dt}\left[\frac{\partial L}{\partial u}\Big(\xi_*(t), \dot{\xi}_*(t), t\Big)\right] = \frac{\partial L}{\partial x}\Big(\xi_*(t), \dot{\xi}_*(t), t\Big)\,,\ \ i=1, \ldots, n\,,} \tag{1.8}$$

writing the Lagrangian as in (1.2). This makes it clear that, to compute the left-hand side of (1.6), we must

1. first evaluate $\frac{\partial L}{\partial \dot{x}}$ "treating $\dot{x}$ as an independent variable,"
2. then plug in the curve $\xi_*(t)$ and its velocity $\dot{\xi}_*(t)$ in the slots for x and $\dot{x}$,
3. finally, differentiate with respect to t.

1.3. The theorem. One possible precise statement of the theorem about the Euler-Lagrange equation is as follows:

THEOREM 1.3.1. *Assume n is a positive integer, Ω is an open subset of $\mathbb{R}^n$, a, b are real numbers such that $a < b$, L is a real-valued function on $\Omega \times \mathbb{R}^n \times [a, b]$, and $\bar{x}$, $\hat{x}$ are given points in Ω. Assume that L is a function of class C^1. Let $\xi_* : [a, b] \mapsto \Omega$ be a curve of class C^1 which is a solution of the minimization problem $\mathcal{P}(L, a, b, \bar{x}, \hat{x})$ in the space of all curves of class C^1. Then the Euler-Lagrange equation (1.8) holds for all $t \in [a, b]$.*

[6]We use the notations that are common today, not Lagrange's. The symbol ∂ for partial derivative was first used by Legendre in 1786.

This is far from being the best possible result. One can prove Theorem 1.3.1 under much weaker hypotheses, but in that case one has to be more careful about the statement. We will discuss such improvements later, and we will postpone the proof until we have become more familiar with the result and with its generalizations.

Notice that, under the hypotheses of Theorem 1.3.1,

1. the curve $t \mapsto (\xi_*(t), \dot{\xi}_*(t), t)$ is continuous,
2. the functions
$$(x, u, t) \mapsto \frac{\partial L}{\partial x}(x, u, t)$$
and
$$(x, u, t) \mapsto \frac{\partial L}{\partial u}(x, u, t)$$
are continuous,

Therefore the conclusion of Theorem 1.3.1 says in fact that the function

$$t \mapsto \frac{\partial L}{\partial u}(\xi_*(t), \dot{\xi}_*(t), t)\,, \tag{1.9}$$

which is only known *a priori* to be continuous, is in fact differentiable everywhere, and its derivative is the function

$$t \mapsto \frac{\partial L}{\partial x}(\xi_*(t), \dot{\xi}_*(t), t)\,. \tag{1.10}$$

Hence *the continuous function (1.9) is in fact continuously differentiable.*

It should be immediately clear from the statement of Theorem 1.3.1 the technical conditions can be weakened. For example:

(A) The statement of Theorem 1.3.1 contains absolutely no reference to any derivatives of L with respect to time. So it ought to be possible to relax the hypothesis that L is of class C^1 and assume instead that L is just differentiable with respect to x and u, maybe with the x- and u-derivatives continuous as functions of (x, u, t).

(B) Suppose we only assume that the curve ξ_* is Lipschitz, and that it is a solution of $\mathcal{P}(L, a, b, \bar{x}, \hat{x})$ in the space of all Lipschitz curves. Then we would of course no longer know *a priori* that the functions (1.9) and (1.10) are continuous, but we would know that they are measurable and bounded. It is not unreasonable to expect that in this case the conclusion of Theorem 1.3.1 might still follow, in the sense that **the function (1.9) is the indefinite integral of the function (1.10)** (that is, equivalently, the function (1.9)—which is only known to be bounded and measurable—is in fact absolutely continuous, and its derivative is equal almost everywhere to the function (1.10).

(C) Suppose we only assume that the curve ξ_* is absolutely continuous, and that it is a solution of $\mathcal{P}(L, a, b, \bar{x}, \hat{x})$ in the space of all absolutely continuous curves. Then we would of course no longer know *a priori* that the functions (1.9) and (1.10) are continuous or bounded, but we would know that they are measurable. Once again, it may not appear unreasonable to expect that in this case the conclusion of Theorem 1.3.1 might still follow, in the sense that **the function (1.9) is the indefinite integral of the function (1.10)** (that is, equivalently, the function (1.9)—which is only known to be and measurable—is in fact absolutely continuous, and its derivative is equal almost everywhere to the function (1.10).

It will turn out that *the conjectures suggested by (A) and (B) are true while, on the other hand, the one suggested by (C) is false, due to the so-called "Lavrentiev phenomenon."*

Moreover, we will see later that Theorem 1.3.1 can be extended even further, in rather surprising ways. For example, *no differentiability of L with respect to the velocity u is needed.* This should of course look strange to you at this point, because if we do not assume that L is differentiable with respect to u then it is not clear at all what the meaning of (1.8) might possibly be. It will turn out, however, that if one interprets things properly, one can get rid of the requirement on differentiability with respect to the velocity u.

On the other hand, it will turn out, also rather surprisingly, that the requirement on differentiability with respect to x can be weakened a bit but not much. This may seem strange because, if you look at (1.8), you may get the impression that more regularity is required for the dependence of L on u than for the dependence of L on x since, after all, in (1.8) we are only differentiating L with respect to x, whereas L is being differentiated with respect to u and then the u-derivative is differentiated again. Yet we will see later that some kind of differentiability with respect to x is crucial no matter what we do, whereas the differentiability with respect to u can be completely avoided if one does things right.

1.4. The momentum. The Euler-Lagrange equation contains the "vector"

$$\boxed{\pi(t) \stackrel{\mathrm{def}}{=} \frac{\partial L}{\partial \dot{x}}(\xi_*(t), \dot{\xi}_*(t), t)\,,} \tag{1.11}$$

known as the *momentum* along the trajectory $\xi_*(\cdot)$. So the components π_i of the momentum are given by

$$\boxed{\pi_i(t) \stackrel{\mathrm{def}}{=} \frac{\partial L}{\partial u^i}(\xi_*(t), \dot{\xi}_*(t), t)\,, \qquad \text{for } i = 1, \ldots, n\,.} \tag{1.12}$$

Naturally, π depends on t, so it is not a "vector" but a *field of "vectors" along the trajectory* ξ_*, that is, a map $t \mapsto \pi(t)$ that assigns to each time t in the interval where ξ_* is defined a "vector" $\pi(t)$. (In modern differential-geometric terminology, π really is a *field of covectors* along $\xi_*(\cdot)$. As we will see later, π shows up in control theory playing an even more important role, and is often referred to by control theorists as the "adjoint vector." This is a somewhat unfortunate choice of terminology, because the $\pi(t)$ are really *covectors* rather than vectors, but it is standard and we will follow it.)

2. COORDINATE INVARIANCE

2.1. Coordinate charts. Now the time has finally come to clarify the distinction between "configurations" and "vectors x in $\mathbb{R}^n$." It is actually quite simple: if we are trying to describe the behavior of a physical system (for example, the motion of a point particle), then this motion will take place in the "configuration space" Q of the system (for example, in three-dimensional physical space for the case of a point particle). Now, *a point of Q is not a tuple of numbers.* We may use tuples of numbers to *represent* points of Q, but these tuples are not the points themselves but, rather, *representations* of the points. Choosing a way to represent points q of Q in a configuration space Q as n-tuples $x = (x^1, \dots, x^n)$ is called *choosing an n-dimensional coordinate chart* on Q. Moreover, this coordinate chart may only be appropriate to represent *some*, but not necessarily *all* the points of Q. This leads us to the following

DEFINITION 2.1.1. Let Q be a set, and let n be a nonnegative integer. An *n-dimensional coordinate chart* on Q is a bijective map

$$D^{\mathbf{x}} \ni q \mapsto \mathbf{x}(q) = (x^1(q), \dots, x^n(q)) \in R^{\mathbf{x}}$$

from some subset $D^{\mathbf{x}}$ (called the *domain* of the chart $\mathbf{x}$) onto an open subset $R^{\mathbf{x}}$ of $\mathbb{R}^n$ (called the *range* of $\mathbf{x}$). $\diamondsuit$

It is important to realize that for a physical system there will always be many different charts. Some of them will be more "natural" than others, as we shall soon see, but even among the "natural" ones there will be no such thing as a canonical choice of *the* chart. For example, even if we want to study the motion of a single particle in three-dimensional physical space, the possible configurations of the particle are the *positions*, that is, points in space, but to represent these positions as triples of numbers we have to choose an origin and three axes (if we are looking for "Cartesian" coordinates), and this can obviously be done in many ways. Moreover, there are also many other, non-Cartesian, ways to represent points, e. g. by means of cylindrical or spherical coordinates.

2.2. Represention of velocity vectors (tangent vectors). Suppose a physical system has configuration space Q, and suppose the system is evolving during a time interval $I \subseteq \mathbb{R}$ in such a way that it describes a curve

$$I \ni t \mapsto \xi(t) \in Q\,.$$

Suppose $\mathbf{x}$ is an n-dimensional chart of our system. Assume. moreover, that the curve ξ is entirely contained in the domain $D^{\mathbf{x}}$ of a chart

$$\mathbf{x} : D^{\mathbf{x}} \mapsto R^{\mathbf{x}} \subseteq \mathbb{R}^n\,.$$

Then our curve ξ admits a *coordinate representation* $\xi^{\mathbf{x}}$ relative to the chart $\mathbf{x}$, given by

$$\xi^{\mathbf{x}}(t) = \mathbf{x}(\xi(t)) \quad \text{for} \quad t \in I\,.$$

This coordinate representation $\xi^{\mathbf{x}}$ is now a curve in $\mathbb{R}^n$, so we may talk, for example, about the curve being "differentiable."

What should we mean by the statement that ξ itself (and not just the coordinate representation $\xi^{\mathbf{x}}$!!!) is "differentiable" at a particular time $t \in I$? One obvious answer is to say that its coordinate representation is differentiable at t. This has the drawback that, as we said before, there are going to be lots of different coordinate charts for Q, which will give rise to different coordinate representations of our curve ξ. But let us postpone the discussion of this difficulty till later, and, temporarily, give the following

Definition 2.2.1. Let Q be a set, and let $I \ni t \mapsto \xi(t) \in Q$ be a Q-valued map defined on an interval $I \subseteq \mathbb{R}$. Let $\mathbf{x}$ be an n-dimensional chart on Q. Assume that $\xi(t) \in D^{\mathbf{x}}$ for all $t \in I$. Let $\bar{t} \in I$. We say that ξ is *differentiable at time* $\bar{t}$ *relative to the chart* $\mathbf{x}$ if the coordinate representation $\xi^{\mathbf{x}}$ is differentiable at $\bar{t}$. $\diamondsuit$

Now we would like to talk about the "velocity" $\dot{\xi}(\bar{t})$ (that is, the derivative of ξ at time $\bar{t}$). Later we will give a precise definition of this object, but even without having done this, it should be reasonable to expect that, whatever $\dot{\xi}(\bar{t})$ turns out to be, it will be represented, relative to the chart $\mathbf{x}$, by the vector $\xi^{\mathbf{x}}(\bar{t})$, which is a perfectly well defined n-tuple of numbers, i. e., a member of $\mathbb{R}^n$.

We can thus give the following

Definition 2.2.2. Let Q, I, ξ, $\mathbf{x}$, $\bar{t}$ be as in Definition 2.2.1, and assume that ξ is differentiable at $\bar{t}$ relative to the chart $\mathbf{x}$. Then the $\mathbf{x}$-*representation of* $\dot{\xi}(\bar{t})$ is the vector $(\dot{\xi}(\bar{t}))^{\mathbf{x}} \in \mathbb{R}^n$ given by

$$(\dot{\xi}(\bar{t}))^{\mathbf{x}} \in \mathbb{R}^n = \dot{\overbrace{\xi^{\mathbf{x}}}}(\bar{t})\,. \qquad \diamondsuit$$

Notice that *we have not yet defined what the "velocity vector"* $\dot{\xi}(\bar{t})$ *is.* We have only said what its representation $(\dot{\xi}(\bar{t}))^{\mathbf{x}}$ relative to a particular chart $\mathbf{x}$ is. Later, we will assign a precise meaning to the expression $\dot{\xi}(\bar{t})$, as a "tangent vector."

2.3. Invariance of the Euler-Lagrange equation. We have now come to the main point of this discussion: the important discovery, made by Lagrange, that *the Euler-Lagrange equation is invariant under arbitrary changes of coordinates.*

Let us first show what this means by doing a straightforward calculation. Suppose we have a curve

$$I \ni t \mapsto \xi_*(t) \in Q$$

and two different coordinate charts

$$\mathbf{x} : D^{\mathbf{x}} \mapsto R^{\mathbf{x}} \subseteq \mathbb{R}^n\,, \qquad \mathbf{X} : D^{\mathbf{X}} \mapsto R^{\mathbf{X}} \subseteq \mathbb{R}^n\,.$$

Suppose our curve ξ_* is contained in both domains, that is,

$$\xi_*(t) \in D^{\mathbf{x}} \cap D^{\mathbf{X}} \quad \text{for} \quad \text{all} \quad t \in I\,.$$

Suppose our two coordinate charts are C^2 *related*, in the following sense:

Definition 2.3.1. Let Q be a set, and let $\mathbf{x}$, $\mathbf{X}$, be two n-dimensional charts on Q. Let k be a nonnegative integer. We say that $\mathbf{x}$ and $\mathbf{X}$ are C^k *related* if

1. the images $\mathbf{x}(D^{\mathbf{x}} \cap D^{\mathbf{X}})$, $\mathbf{X}(D^{\mathbf{x}} \cap D^{\mathbf{X}})$ of the "overlap set" $D^{\mathbf{x}} \cap D^{\mathbf{X}}$ under the coordinate maps $\mathbf{x}$, $\mathbf{X}$, are open in $R^{\mathbf{x}}$, $R^{\mathbf{X}}$, respectively;

2. the "change of coordinates" maps

$$\Phi^{\mathbf{x},\mathbf{X}} : \mathbf{x}(D^{\mathbf{x}} \cap D^{\mathbf{X}}) \mapsto \mathbf{X}(D^{\mathbf{x}} \cap D^{\mathbf{X}}),$$
$$\Phi^{\mathbf{X},\mathbf{x}} : \mathbf{X}(D^{\mathbf{x}} \cap D^{\mathbf{X}}) \mapsto \mathbf{x}(D^{\mathbf{x}} \cap D^{\mathbf{X}}),$$

defined by the conditions

$$\Phi^{\mathbf{x},\mathbf{X}}(x) = X \quad \text{whenever} \quad x = \mathbf{x}(q)\,,\ X = \mathbf{X}(q) \quad \text{for some} \quad q \in D^{\mathbf{x}} \cap D^{\mathbf{X}}\,,$$
$$\Phi^{\mathbf{X},\mathbf{x}}(X) = x \quad \text{whenever} \quad x = \mathbf{x}(q)\,,\ X = \mathbf{X}(q) \quad \text{for some} \quad q \in D^{\mathbf{x}} \cap D^{\mathbf{X}}\,,$$

are of class C^k. ♢

(The map $\Phi^{\mathbf{x},\mathbf{X}}$ is the "coordinate change from $\mathbf{x}$ to $\mathbf{X}$," that is, the map that assigns, to each point $x \in R^{\mathbf{x}}$ which is the $\mathbf{x}$-coordinate representation of some $q \in D^{\mathbf{x}} \cap D^{\mathbf{X}}$, the point X that represents the same q in the $\mathbf{X}$-chart. It is clear that $\Phi^{\mathbf{x},\mathbf{X}}$ is well defined, because, if $x \in R^{\mathbf{x}}$ is the $\mathbf{x}$-coordinate representation of some $q \in D^{\mathbf{x}} \cap D^{\mathbf{X}}$, then this q is unique, because $\mathbf{x}$ is injective, and $\mathbf{X}(q)$ exists, because $q \in D^{\mathbf{X}}$. Naturally, similar remarks apply to the map $\Phi^{\mathbf{X},\mathbf{x}}$.)

Now suppose L is a "Lagrangian on Q." The precise meaning of this is that L is a function of points $q \in Q$, "velocity vectors" (i. e., tangent vectors) $\dot q$, and time, that is, L is a function

$$Q|\dot Q \times [a,b] \ni (q,\dot q,t) \mapsto L(q,\dot q,t) \in \mathbb{R}\,,$$

defined at points $(q,\dot q)$ in the "configuration-velocity space" $Q|\dot Q$ and times t in some interval $[a,b]$. More precisely, what we are now temporarily calling "configuration-velocity space" will later be called "the tangent bundle of configuration space," and the mysterious "velocity vectors" will be interpreted as "tangent vectors." Since we have not yet defined the notion of a tangent vector, we shall for the time being assume that L—whatever it may be—is specified by giving its representation $(x,\dot x,t) \mapsto L^{\mathbf{x}}(x,\dot x,t)$, for each coordinate chart $\mathbf{x}$, by means of a function $L^{\mathbf{x}}$.

In addition, we shall assume that "the values of $L^{\mathbf{x}}(x,\dot x,t)$ and $L^{\mathbf{X}}(X,\dot X,t)$ are the same whenever x, X are coordinate representations of the same point q, and $\dot x$, $\dot X$ are coordinate representations of the same 'velocity' $\dot q$ relative to two charts $\mathbf{x}$, $\mathbf{X}$." First of all, we have to decide what this means. Clearly, what it should mean is that

$$L^{\mathbf{x}}(\hat{\mathbf{x}}(q,\dot q),t) = L^{\mathbf{X}}(\hat{\mathbf{X}}(q,\dot q),t) \quad \text{whenever} \quad (q,\dot q) \in D^{\hat{\mathbf{x}}} \cap D^{\hat{\mathbf{X}}}\,, \tag{2.1}$$

where $\hat{\mathbf{x}}(q,\dot q)$, $\hat{\mathbf{X}}(q,\dot q)$, are the coordinate representations of a "configuration-velocity point" $(q,\dot q)$ with respect to the charts $\mathbf{x}$, $\mathbf{X}$. Equivalently, (2.1) should say that

$$\begin{aligned} &L^{\mathbf{x}}(\hat x,t) = L^{\mathbf{X}}(\hat X,t) \quad \text{whenever} \\ &\hat x = \hat{\mathbf{x}}(q,\dot q) \text{ and } \hat X = \hat{\mathbf{X}}(q,\dot q) \ \text{ for some } \in D^{\hat{\mathbf{x}}} \cap D^{\hat{\mathbf{X}}}\,. \end{aligned} \tag{2.2}$$

The first thing we have to do is figure out what (2.2) means directly in terms of the functions $L^{\mathbf{x}}$ and $L^{\mathbf{X}}$, without going through the intermediate point $(q,\dot q)$. Clearly, the vectors $\hat x$, $\hat X$ are coordinate representations of a position and velocity, so they should belong to $\mathbb{R}^{2n}$ and be given as pairs $(x,\dot x)$, $(X,\dot X)$, where x, X represent q and $\dot x, \dot X$ represent $\dot q$. Moreover, as long as $q \in D^{\hat{\mathbf{x}}} \cap D^{\hat{\mathbf{X}}}$, we already know how to express X in terms of x, using the map $\Phi^{\mathbf{x},\mathbf{X}}$, and there ought to be a similar way to "express $\dot x$ in terms of $\dot X$." (As we will soon see, this is not 100% accurate. We will express $\dot x$ in terms of X and $\dot X$, not just in terms of $\dot X$ alone.) Let us find the formula for this. Using Φ for $\Phi^{\mathbf{x},\mathbf{X}}$, and Ψ for $\Phi^{\mathbf{X},\mathbf{x}}$, we have

$$\begin{aligned} X^i &= \Phi^i(x^1,\ldots,x^n) \\ x^j &= \Psi^j(X^1,\ldots,X^n)\,, \end{aligned}$$

for $i = 1,\ldots,n$, $j = 1,\ldots,n$. Then we can express $\dot{X}$ as a function $\check{\Phi}(x,\dot{x})$, and $\dot{x}$ as a function $\check{\Psi}(X,\dot{X})$, by differentiating. The result is

$$\dot{X}^i = \check{\Phi}^i(x^1,\ldots,x^n,\dot{x}^1,\ldots,\dot{x}^n) = \sum_{j=1}^n \frac{\partial \Phi^i}{\partial x^j}(x^1,\ldots,x^n)\dot{x}^j\,, \tag{2.3}$$

$$\dot{x}^j = \check{\Psi}^j(X^1,\ldots,X^n,\dot{X}^1,\ldots,\dot{X}^n) = \sum_{i=1}^n \frac{\partial \Psi^j}{\partial X^i}(X^1,\ldots,X^n)\dot{X}^i\,, \tag{2.4}$$

for $i = 1,\ldots,n$, $j = 1,\ldots,n$. So now we are able to write the complete formulae relating $\hat{x} = (x,\dot{x})$ and $\hat{X} = (X,\dot{X})$:

$$X^i = \Phi^i(x^1,\ldots,x^n)\,, \tag{2.5}$$

$$\dot{X}^i = \check{\Phi}^i(x^1,\ldots,x^n,\dot{x}^1,\ldots,\dot{x}^n)\,, \tag{2.6}$$

$$x^j = \Psi^j(X^1,\ldots,X^n)\,, \tag{2.7}$$

$$\dot{x}^j = \check{\Psi}^j(X^1,\ldots,X^n,\dot{X}^1,\ldots,\dot{X}^n)\,, \tag{2.8}$$

for $i = 1,\ldots,n$, $j = 1,\ldots,n$, where the functions $\check{\Phi}^i$, $\check{\Psi}^j$ are defined by (2.3) and (2.4). (Notice that, as we announced before, we cannot express $\dot{X}$ just in terms of $\dot{x}$; the formulas actually give $\dot{X}$ as a function of x and $\dot{x}$.)

The identities (2.3) and (2.4) are the *transformation formulas for velocity vectors*. They tell us how to find the components of such a vector in one coordinate system if we know the components in another coordinate system.

Let us now compute the partial derivatives of the functions given by (2.5), (2.6), (2.7), (2.8). We have

$$\begin{aligned}
\frac{\partial X^i}{\partial x^j} &= \frac{\partial \Phi^i}{\partial x^j}(x^1,\ldots,x^n)\,, \\
\frac{\partial \dot{X}^i}{\partial x^j} &= \sum_{k=1}^n \frac{\partial^2 \Phi^i}{\partial x^j \partial x^k}(x^1,\ldots,x^n)\dot{x}^k\,, \\
\frac{\partial \dot{X}^i}{\partial \dot{x}^j} &= \frac{\partial \Phi^i}{\partial x^j}(x^1,\ldots,x^n)\,, \\
\frac{\partial x^j}{\partial X^i} &= \frac{\partial \Psi^j}{\partial X^i}(X^1,\ldots,X^n)\,, \\
\frac{\partial \dot{x}^j}{\partial X^i} &= \sum_{k=1}^n \frac{\partial^2 \Psi^j}{\partial X^i \partial X^k}(X^1,\ldots,X^n)\dot{X}^k\,, \\
\frac{\partial \dot{x}^j}{\partial \dot{X}^i} &= \frac{\partial \Psi^j}{\partial X^i}(X^1,\ldots,X^n)\,.
\end{aligned}$$

Now let us study how the components of the momentum are represented in our two charts $\mathbf{x}$, $\mathbf{X}$. Let us use p, P to denote the two representations, so

$$p = (p_1,\ldots,p_n) = \Big(\frac{\partial L}{\partial \dot{x}^1},\ldots,\frac{\partial L}{\partial \dot{x}^n}\Big)$$

and

$$P = (P_1, \dots, P_n) = \Big(\frac{\partial L}{\partial \dot X^1}, \dots, \frac{\partial L}{\partial \dot X^n} \Big) .$$

Repeated use of the Chain Rule gives

$$\begin{aligned}
\frac{\partial L}{\partial X^i} &= \sum_{k=1}^n \frac{\partial L}{\partial x^k} \cdot \frac{\partial x^k}{\partial X^i} + \sum_{k=1}^n \frac{\partial L}{\partial \dot x^k} \cdot \frac{\partial \dot x^k}{\partial X^i} \\
&= \sum_{k=1}^n \frac{\partial L}{\partial x^k} \cdot \frac{\partial \Psi^k}{\partial X^i} + \sum_{k=1}^n \sum_{\ell=1}^n \frac{\partial L}{\partial \dot x^k} \cdot \frac{\partial^2 \psi^k}{\partial X^i \partial X^\ell} \dot X^\ell , \\
P_i &= \frac{\partial L}{\partial \dot X^i} \\
&= \sum_{j=1}^n \frac{\partial L}{\partial x^j} \cdot \frac{\partial x^j}{\partial \dot X^i} + \sum_{j=1}^n \frac{\partial L}{\partial \dot x^j} \cdot \frac{\partial \dot x^j}{\partial \dot X^i} \\
&= \sum_{j=1}^n \frac{\partial L}{\partial \dot x^j} \cdot \frac{\partial \dot x^j}{\partial \dot X^i} \\
&= \sum_{j=1}^n \frac{\partial L}{\partial \dot x^j} \cdot \frac{\partial \Psi^j}{\partial X^i} \\
&= \sum_{j=1}^n p_j \cdot \frac{\partial \Psi^j}{\partial X^i} .
\end{aligned}$$

So we got the formula

$$P_i = \sum_{j=1}^n p_j \cdot \frac{\partial \Psi^j}{\partial X^i} , \tag{2.9}$$

and an identical argument also yields the identity

$$p_j = \sum_{i=1}^n P_i \cdot \frac{\partial \Phi^i}{\partial x^j} . \tag{2.10}$$

Let us look at these transformation formulae together with the identities we derived earlier for velocity vectors:

$$P_i = \sum_{j=1}^n \frac{\partial \Psi^j}{\partial X^i} p_j , \tag{2.11}$$

$$p_j = \sum_{i=1}^n \frac{\partial \Phi^i}{\partial x^j} P_i , \tag{2.12}$$

$$\dot X^i = \sum_{j=1}^n \frac{\partial \Phi^i}{\partial x^j} \dot x^j , \tag{2.13}$$

$$\dot x^j = \sum_{i=1}^n \frac{\partial \Psi^j}{\partial X^i} \dot X^i . \tag{2.14}$$

Notice that *the transformation formulae for the momentum are different from the formulae for the velocities.* The formula giving you $\dot X$ in terms of $\dot x$ is the same as that giving you p in terms of P. In other ways, *the momentum and the velocity* *transform in opposite ways.* The same rule that enables you to go from $\dot{x}$ to $\dot{X}$ will take you from P to p, not from p to P!!

For historical reasons, objects such as the momentum are called *covariant vectors*, or, simply, *covectors*. Objects such as the velocity, which transform in the opposite way, are called *contravariant vectors*, or, simply, *vectors*, or *ordinary vectors*, or *tangent vectors*.

REMARK 2.3.2. You may have noticed that we used *superscripts* for the components of velocity vectors, and *subscripts* for the components of momentum "vectors." This is common practice in differential geometry and physics, and there are very good reasons for it, which will be discussed later. ♢

We now return to the issue of the invariance of the Euler-Lagrange equation. The equation says that a certain "vector" is equal to zero. This "Euler-Lagrange vector" is the one having components

$$e_i = a_i - b_i$$

relative to the $\mathbf{x}$ chart, where

$$\begin{aligned} a_i &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\right), \\ b_i &= \frac{\partial L}{\partial x^i}. \end{aligned}$$

In the $\mathbf{X}$ system, the ""Euler-Lagrange vector" will have components

$$E_i = A_i - B_i$$

where

$$\begin{aligned} A_i &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{X}^i}\right), \\ B_i &= \frac{\partial L}{\partial X^i}. \end{aligned}$$

We now have to find the transformation formula relating the E_i to the e_i. For this purpose, we first study the transformation of the a_i to the A_i and of the b_i to the B_i. We have

$$\begin{aligned} A_i &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{X}^i}\right) \\ &= \sum_{k=1}^{n} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right) \cdot \frac{\partial \Psi^k}{\partial X^i} + \sum_{k=1}^{n} \frac{\partial L}{\partial \dot{x}^k} \cdot \frac{d}{dt}\left(\frac{\partial \Psi^k}{\partial X^i}\right) \\ &= \sum_{k=1}^{n} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right) \cdot \frac{\partial \Psi^k}{\partial X^i} + \sum_{k=1}^{n}\sum_{\ell=1}^{n} \frac{\partial L}{\partial \dot{x}^k} \cdot \frac{\partial^2 \Psi^k}{\partial X^i \partial X^\ell} \dot{X}^\ell \\ &= \sum_{k=1}^{n} a_k \cdot \frac{\partial \Psi^k}{\partial X^i} + \sum_{k=1}^{n}\sum_{\ell=1}^{n} p_k \cdot \frac{\partial^2 \Psi^k}{\partial X^i \partial X^\ell} \dot{X}^\ell, \end{aligned}$$

$$\begin{aligned}
B_i &= \frac{\partial L}{\partial X^i} \\
&= \sum_{k=1}^{n} \frac{\partial L}{\partial x^k} \cdot \frac{\partial \Psi^k}{\partial X^i} + \sum_{k=1}^{n}\sum_{\ell=1}^{n} \frac{\partial L}{\partial \dot{x}^k} \cdot \frac{\partial^2 \psi^k}{\partial X^i \partial X^\ell} \dot{X}^\ell \\
&= \sum_{k=1}^{n} b_k \cdot \frac{\partial \Psi^k}{\partial X^i} + \sum_{k=1}^{n}\sum_{\ell=1}^{n} p_k \cdot \frac{\partial^2 \psi^k}{\partial X^i \partial X^\ell} \dot{X}^\ell \,,
\end{aligned}$$

and then

$$\begin{aligned}
E_i &= A_i - B_i \\
&= \sum_{k=1}^{n} a_k \cdot \frac{\partial \Psi^k}{\partial X^i} + \sum_{k=1}^{n}\sum_{\ell=1}^{n} p_k \cdot \frac{\partial^2 \Psi^k}{\partial X^i \partial X^\ell} \dot{X}^\ell \\
&\quad - \sum_{k=1}^{n} b_k \cdot \frac{\partial \Psi^k}{\partial X^i} - \sum_{k=1}^{n}\sum_{\ell=1}^{n} p_k \cdot \frac{\partial^2 \Psi^k}{\partial X^i \partial X^\ell} \dot{X}^\ell \\
&= \sum_{k=1}^{n} a_k \cdot \frac{\partial \Psi^k}{\partial X^i} - \sum_{k=1}^{n} b_k \cdot \frac{\partial \Psi^k}{\partial X^i} \\
&= \sum_{k=1}^{n} (a_k - b_k) \cdot \frac{\partial \Psi^k}{\partial X^i} \\
&= \sum_{k=1}^{n} e_k \cdot \frac{\partial \Psi^k}{\partial X^i} \,.
\end{aligned}$$

So we have derive the formula

$$E_i = \sum_{k=1}^{n} e_k \cdot \frac{\partial \Psi^k}{\partial X^i} \,, \tag{2.15}$$

from which it follows that if all the the e_i vanish then all the E_i vanish as well. An identical argument works in the other direction as well, so we have in fact shown that

$$e_1 = e_2 = \ldots = e_n = 0 \quad \Longleftrightarrow \quad E_1 = E_2 = \ldots = E_n = 0 \,. \tag{2.16}$$

In other words

The Euler-Lagrange equation holds in the **x** *chart if and only if it holds in the* **X** *chart.*

This is the desired invariance statement.

2.4. A comment on vectors, covectors, and tensors. Let us compare the transformation formulae for the velocity, the momentum, the "Euler-Lagrange vector," and other objects we have introduced, such as the two parts whose difference is the Euler-Lagrange vector. The formulas say that

$$\dot{X}^i = \sum_{k=1}^{n} \frac{\partial \Phi^i}{\partial x^k} \dot{x}^k \,, \tag{2.17}$$

$$\dot{x}^j = \sum_{k=1}^{n} \frac{\partial \Psi^j}{\partial X^k} \dot{X}^k \,, \tag{2.18}$$

$$P_i = \sum_{k=1}^{n} \frac{\partial \Psi^k}{\partial X^i} p_k \,, \tag{2.19}$$

$$p_j = \sum_{k=1}^{n} \frac{\partial \Phi^k}{\partial x^j} P_k \,, \tag{2.20}$$

$$E_i = \sum_{k=1}^{n} \frac{\partial \Psi^k}{\partial X^i} e_k \,, \tag{2.21}$$

$$e_j = \sum_{k=1}^{n} \frac{\partial \Phi^k}{\partial x^j} E_k \,. \tag{2.22}$$

$$A_i = \sum_{k=1}^{n} a_k \cdot \frac{\partial \Psi^k}{\partial X^i} + \sum_{k=1}^{n}\sum_{\ell=1}^{n} p_k \cdot \frac{\partial^2 \Psi^k}{\partial X^i \partial X^\ell} \dot{X}^\ell \,, \tag{2.23}$$

$$a_j = \sum_{k=1}^{n} A_k \cdot \frac{\partial \Phi^k}{\partial x^i} + \sum_{k=1}^{n}\sum_{\ell=1}^{n} P_k \cdot \frac{\partial^2 \Phi^k}{\partial x^i \partial x^\ell} \dot{x}^\ell \,, \tag{2.24}$$

$$B_i = \sum_{k=1}^{n} b_k \cdot \frac{\partial \Psi^k}{\partial X^i} + \sum_{k=1}^{n}\sum_{\ell=1}^{n} p_k \cdot \frac{\partial^2 \psi^k}{\partial X^j \partial X^\ell} \dot{X}^\ell \,. \,, \tag{2.25}$$

$$b_j = \sum_{k=1}^{n} B_k \cdot \frac{\partial \Phi^k}{\partial x^j} + \sum_{k=1}^{n}\sum_{\ell=1}^{n} P_k \cdot \frac{\partial^2 \Phi^k}{\partial X^i \partial x^\ell} \dot{x}^\ell \,. \tag{2.26}$$

These formulas show that:

1. The transformation law for the es and Es is the same as that for the ps and Ps. In other words: *the Euler-Lagrange "vector" is really a covector, like the momentum.*
2. The transformation laws for the xs and Xs are "opposite" from those for the ps and Ps. This why velocities are called *contravariant vectors.*
3. In all three cases (that is, in the transformation laws for the es and Es, the ps and Ps, and the xs and Xs), the laws are such that *if the object under consideration vanishes in one coordinate chart then it vanishes in all of them.* That is,
   (T) *the vanishing of the object is an invariant property, not depending on the choice of a coordinate chart.*

   Later, we will define other quantities, called *tensors*, which are more general than vectors and covectors, but still have Property (T). Property (T) will turn out to be a manifestation of the *tensor nature* of the velocity and the momentum.
4. The transformation laws for the as and As and the bs and Bs are quite different. In particular, *Property (T) does not hold.* So *the two parts whose difference is the Euler-Lagrange covector are not themselves covectors, and are not even tensors of any kind.*