Chapter 3

Second-Order ODE and the Calculus of Variations

3.1. Tangent Vectors and the Tangent Bundle

Let $\sigma : I \to \mathbf{R}^n$ be a C^1 curve in $\mathbf{R}^n$ and suppose that $\sigma(t_0) = p$ and $\sigma'(t_0) = v$. Up until this point we have referred to v as either the velocity or the tangent vector to σ at time t_0. From now on we will make a small but important distinction between these two concepts. While the distinction is not critical in dealing with first-order ODE, it will simplify our discussion of second-order ODE. Henceforth we will refer to v as the velocity of σ at time t_0 and define its tangent vector at time t_0 to be the ordered pair $\dot{\sigma}(t_0) := (p, v)$. (If you are familiar with the distinction that is sometimes made between "free" and "based" vectors, you will recognize this as a special case of that.)

The set of all tangent vectors (p, v) we get in this way is called the *tangent bundle* of $\mathbf{R}^n$, and we will denote it by $\mathbf{TR}^n$. Clearly $\mathbf{TR}^n = \mathbf{R}^n \times \mathbf{R}^n$, so you might think that it is a useless complication to introduce this new symbol, but in fact there is a good reason for using this kind of redundant notation. As mathematical constructs get more complex, it is important to have notation that gives visual clues about what symbols mean. So, in particular, when we want to emphasize that a pair (p, v) is referring to a tangent vector, it is better to write $(p, v) \in \mathbf{TR}^n$ rather than $(p, v) \in \mathbf{R}^n \times \mathbf{R}^n$.

The projection of $\mathbf{TR}^n$ onto its first factor, taking (p, v) to its base point p, will be denoted by $\Pi : \mathbf{TR}^n \to \mathbf{R}^n$, and it is called the tangent bundle projection. (There is of course another projection, onto the second factor, but it will play no role in what follows and we will not give it a name.) The set of all tangent vectors that project onto a fixed base point p will be denoted by $\mathbf{T}_p\mathbf{R}^n$. It is called the tangent space to $\mathbf{R}^n$ at the point p, and its elements are called tangent vectors at p. It is clearly an n-dimensional real vector space, and its dual space, the space of linear maps of $\mathbf{T}_p\mathbf{R}^n$ into $\mathbf{R}$, is called the cotangent space of $\mathbf{R}^n$ at p and is denoted by $\mathbf{T}^*_p\mathbf{R}^n$.

Let f be a smooth real-valued function defined in an open set O of $\mathbf{R}^n$. If $p \in O$ and $V = (p, v) \in \mathbf{T}_p\mathbf{R}^n$, recall that Vf, the directional derivative of f at p in the direction v, is defined as $Vf := \sum_{i=1}^n v_i \frac{\partial f}{\partial x_i}(p)$ (see Appendix C). If as above $\sigma : I \to \mathbf{R}^n$ is a smooth curve and $V = \dot\sigma(t_0)$ is its tangent vector at time t_0, then by the chain rule $\left(\frac{d}{dt}\right)_{t=t_0} f(\sigma(t)) = Vf$. We will follow the customary practice of using the symbolic differentiation operator $\sum_{i=1}^n v_i (\frac{\partial}{\partial x_i})_p$ as an alternative notation for the tangent vector V. If we fix f, then $V \mapsto Vf$ is a linear functional, df_p, on $\mathbf{T}_p\mathbf{R}^n$ (i.e., an element of $\mathbf{T}^*_p\mathbf{R}^n$) called the differential of f at p.

Note that the function f defined on O gives rise to two associated functions in $\Pi^{-1}(O)$. The first is just $f \circ \Pi$, $(p, v) \mapsto f(p)$, and the second is df, $(p, v) \mapsto df_p(v)$. This last remark is the basis of a very important construction that "promotes" a system of local coordinates $(x_1, \dots, x_n)$ for $\mathbf{R}^n$ in O (see Appendix D) to a system of local coordinates $(q_1, \dots, q_n, \dot q_1, \dots, \dot q_n)$ for $\mathbf{TR}^n$ in $\Pi^{-1}(O)$ called the associated *canonical coordinates* for the tangent bundle. Namely, we define $q_i := x_i \circ \Pi$ and $\dot q_i := dx_i$. Suppose $V_1 = (p_1, v_1)$ and $V_2 = (p_2, v_2)$ are in $\Pi^{-1}(O)$ and that $q_i(V_1) = q_i(V_2)$ and $\dot q_i(V_1) = \dot q_i(V_2)$ for all $i = 1, \dots, n$. Since the x_i are local coordinates in O and $x_i(p_1) = q_i(V_1) = q_i(V_2) = x_i(p_2)$, it follows that $p_1 = p_2 = p$. It also follows from the fact that x_i is a local coordinate system that the $(dx_i)_p$ are a basis for $\mathbf{T}_p\mathbf{R}^n$; hence $(dx_i)_p(v_1) = \dot q_i(V_1) = \dot q_i(V_2) = (dx_i)_p(v_2)$ implies $v_1 = v_2$. Thus $V_1 = V_2$, so $(q_1, \dots, q_n, \dot q_1, \dots, \dot q_n)$ really are local coordinates for $\mathbf{TR}^n$ in $\Pi^{-1}(O)$. Note that for each p in O, $(\dot{q}_1, \dots, \dot{q}_n)$ are actually Cartesian coordinates on $\mathbf{T}_p\mathbf{R}^n$. We will be using canonical coordinates almost continuously from now on.

▷ **Exercise 3–1.** Check that if $(x_1, \dots, x_n)$ are the standard coordinates for $\mathbf{R}^n$, then $(q_1, \dots, q_n, \dot{q}_1, \dots, \dot{q}_n)$ are the standard coordinates for $\mathbf{TR}^n = \mathbf{R}^{2n}$.

In Appendix C we define a certain natural vector field R on $\mathbf{R}^n$—called the radial or Euler vector field—by $R(x) := x$, and we point out that it has the same expression in every Cartesian coordinate system $(x_1, \dots, x_n)$, namely $R = \sum_{i=1}^n x_i \frac{\partial}{\partial x_i}$. We also recall there Euler's famous theorem on homogeneous functions, which states that if $f : \mathbf{R}^n \to \mathbf{R}$ is C^1 and is positively homogeneous of degree k (meaning $f(tx) = t^k f(x)$ for all $t > 0$ and $x \neq 0$), then $Rf = \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i} = kf$. We now define a vector field R on $\mathbf{TR}^n$, also called the radial or Euler vector field, by defining it on each tangent space $\mathbf{T}_p\mathbf{R}^n$ to be the radial vector field on $\mathbf{T}_p\mathbf{R}^n$. Explicitly, $R(p, v) = (0, v)$. Recalling that $(\dot{q}_1, \dots, \dot{q}_n)$ are Cartesian coordinates on $\mathbf{T}_p\mathbf{R}^n$ we have

3.1.1. Proposition. *If $(q_1, \dots, q_n, \dot{q}_1, \dots, \dot{q}_n)$ are canonical coordinates for $\mathbf{TR}^n$ in $\Pi^{-1}(O)$, then the radial vector field R on $\mathbf{TR}^n$ has the expression*

$$R = \sum_{i=1}^n \dot{q}_i \frac{\partial}{\partial \dot{q}_i}$$

in $\Pi^{-1}(O)$. Hence if $F : \mathbf{TR}^n \to \mathbf{R}$ is a C^1 real-valued function that is positively homogeneous of degree k on each tangent space $\mathbf{T}_p\mathbf{R}^n$, then $RF = \sum_{i=1}^n \dot{q}_i \frac{\partial F}{\partial \dot{q}_i} = kF$.

Suppose that $\sigma : I \to \mathbf{R}^n$ is a C^1 curve. A path $\tilde{\sigma} : I \to \mathbf{TR}^n$ is called a *lifting* of σ if it projects onto σ under Π, i.e., if it is of the form $\tilde{\sigma}(t) = (\sigma(t), \lambda(t))$ for some C^1 map $\lambda : I \to \mathbf{R}^n$. You should think of a lifting of σ as being a vector field defined along σ. There are many different possible liftings of σ. For example, the lifting $t \mapsto (\sigma(t), 0)$ is the zero vector field along σ, and $t \mapsto (\sigma(t), \sigma''(t))$ is the acceleration vector field of σ. The tangent vector field $\dot{\sigma}$, $t \mapsto (\sigma(t), \sigma'(t))$ plays an especially important role, and we shall also refer to it as the *canonical* *lifting* of σ to the tangent bundle. We note that it takes C^k curves in $\mathbf{R}^n$ $(k \geq 1)$ to C^{k-1} curves in $\mathbf{TR}^n$.

Let $x^1, \ldots, x^n$ be local coordinates in O, and assume σ maps I into O. In these coordinates the curve σ is described by its so-called parametric representation, $x_i(t) := x^i(\sigma(t))$. Let's see what the parametric representation is for the canonical lifting, $\dot{\sigma}(t)$, with respect to the canonical coordinates $(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n)$ defined by the x^i. Since $\Pi \circ \dot{\sigma}(t) = \sigma(t)$, it follows from the definition of the q_i that

$$q_i(\dot{\sigma}(t)) = x^i(\Pi(\dot{\sigma}(t))) = x_i(t).$$

On the other hand,

$$\dot{q}_i(\dot{\sigma}(t)) = dx^i(\sigma'(t)) = \frac{d}{dt}x^i(\sigma(t)) = \frac{dx_i(t)}{dt}.$$

3.2. Second-Order Differential Equations

We have seen that solving a first-order differential equation in $\mathbf{R}^n$ involves finding a path $x(t)$ in $\mathbf{R}^n$ given

1) its initial position and
2) its velocity as a function of its position and the time.

Similarly, solving a second-order differential equation in $\mathbf{R}^n$ involves finding a path $x(t)$ in $\mathbf{R}^n$ given

1) its initial position **and** its initial velocity and
2) its acceleration as a function of its position, its velocity, and the time.

To make this precise, suppose $A : \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}^n$ is a C^1 function. A C^2 curve $x(t)$ in $\mathbf{R}^n$ is said to be a solution of the second-order differential equation in $\mathbf{R}^n$, $\frac{d^2x}{dt^2} = A\left(x, \frac{dx}{dt}, t\right)$, if $x''(t) = A(x(t), x'(t), t)$ holds for all t in the domain of x. Given such a second-order differential equation on $\mathbf{R}^n$, the associated initial value problem (or IVP) is to find a solution $x(t)$ for which both the position $x(t_0)$ and the velocity $x'(t_0)$ have some specified values at a particular time t_0.

Fortunately, we do not have to start all over from scratch to develop theory and intuition concerning second-order equations. There is an easy trick that we already remarked on in Chapter 1 that effectively reduces the consideration of a second-order differential equation in $\mathbf{R}^n$ to the consideration of a first-order equation in $\mathbf{TR}^n = \mathbf{R}^n \times \mathbf{R}^n$. Namely, given $A : \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}^n$ as above, define a time-dependent vector field V on $\mathbf{TR}^n$ by $V(p, v, t) = (v, A(p, v, t))$. Suppose first that $(x(t), v(t))$ is a C^1 path in $\mathbf{TR}^n$. Note that its derivative is just $(x'(t), v'(t))$, so this path is a solution of the first-order differential equation $\frac{d(x,v)}{dt} = V(x, v, t)$ if and only if $(x'(t), v'(t)) = V(x(t), v(t), t) = (v(t), A(x(t), v(t), t))$, which of course means that $x'(t) = v(t)$ while $v'(t) = A(x(t), v(t), t)$. But then $x''(t) = v'(t) = A(x(t), v(t), t)$, so $x(t)$ is a solution of the second-order equation $\frac{d^2x}{dt^2} = A\left(x, \frac{dx}{dt}, t\right)$. Conversely, if $x(t)$ is a solution of $\frac{d^2x}{dt^2} = A\left(x, \frac{dx}{dt}, t\right)$ and we define $v(t) = x'(t)$, then $(x(t), v(t))$ is the canonical lifting of $x(t)$ and is a solution of $\frac{d(x,v)}{dt} = V(x, v, t)$. What we have shown is

3.2.1. Reduction Theorem for Second-Order ODE. *The canonical lifting of C^2 curves in $\mathbf{R}^n$ to C^1 curves on $\mathbf{TR}^n$ sets up a bijective correspondence between solutions of the second-order differential equation $\frac{d^2x}{dt^2} = A\left(x, \frac{dx}{dt}, t\right)$ on $\mathbf{R}^n$ and solutions of the first-order equation $\frac{d(x,v)}{dt} = V(x, v, t)$ on $\mathbf{TR}^n$, where $V(x(t), v(t), t) = (v(t), A(x(t), v(t), t))$.*

▷ **Exercise 3–2.** Use this correspondence to formulate and prove existence, uniqueness, and smoothness theorems for second-order differential equations from the corresponding theorems for first-order differential equations. Extend this to higher-order differential equations.

According to the Reduction Theorem, a second-order ODE in $\mathbf{R}^n$ is described by some vector field on $\mathbf{TR}^n$. But be careful, not every vector field V on $\mathbf{TR}^n$ arises in this way.

▷ **Exercise 3–3.** Show that a time-dependent vector field $V(p, v, t)$ on $\mathbf{TR}^n$ arises as above from some second-order ODE on $\mathbf{R}^n$, $\frac{d^2x}{dt^2} = A\left(x, \frac{dx}{dt}, t\right)$, if and only if $\Pi(V(p, v, t)) = v$ for all $(p, v, t) \in \mathbf{TR}^n \times \mathbf{R}$.

▷ **Exercise 3–4.** Let $x^1, \dots, x^n$ be local coordinates in some open set O of $\mathbf{R}^n$ and let $q_1, \dots, q_n, \dot{q}_1, \dots, \dot{q}_n$ be the associated canonical coordinates in $\mathbf{TR}^n$. Suppose that $\sigma : I \to O$ is a smooth path and $x_i(t) := x^i(\sigma(t))$ its parametric representation, and let $q_i(t) := q_i(\dot{\sigma}(t))$, $\dot{q}_i(t) := \dot{q}_i(\dot{\sigma}(t))$ be the parametric representation of the canonical lifting $\dot{\sigma}$. Show that σ is a solution of the second-order ODE $\frac{d^2x}{dt^2} = A\left(x, \frac{dx}{dt}, t\right)$ if and only if for all $t \in I$, $\frac{d\dot{q}_i(t)}{dt} = A_i(q_1(t), \dots, q_n(t), \dot{q}_1(t), \dots, \dot{q}_n(t), t)$. (Hint: Recall that $q_i(t) := x_i(t)$ and $\dot{q}_i(t) := \frac{dx_i(t)}{dt}$.)

3.2.2. Definition. Let $\frac{d^2x}{dt^2} = A\left(x, \frac{dx}{dt}\right)$ be a time-independent second-order ODE on $\mathbf{R}^n$. A function $f : \mathbf{TR}^n \to \mathbf{R}$ is called a *constant of the motion* (or a *conserved quantity* or a *first integral*) of this ODE if f is constant along the canonical lifting, $\dot{\sigma}$, of every solution curve σ (equivalently, if whenever $x(t)$ satisfies the ODE, then $f(x(t), x'(t))$ is a constant).

▷ **Exercise 3–5.** Let V be a vector field in $\mathbf{TR}^n$ defined by $V(p, v) := (v, A(p, v))$. Show that $f : \mathbf{TR}^n \to \mathbf{R}$ is a constant of the motion of $\frac{d^2x}{dt^2} = A\left(x, \frac{dx}{dt}\right)$ if and only if the directional derivative, Vf, of f in the direction V, is identically zero.

3.3. The Calculus of Variations

Where do second-order ordinary differential equations come from—or to phrase this question somewhat differently, what sort of things get represented mathematically as solutions of second-order ODEs?

Perhaps the first answer that will spring to mind for many people is Newton's Second Law of Motion, "$F = ma$", which without doubt inspired much of the early work on second-order ODE. But as we shall soon see, important as Newton's Equations of Motion are, they are best seen mathematically as a special case of a much more general class of second-order ODE, called Euler-Lagrange Equations, and a more satisfying answer to our question will grow out of an understanding of this family of equations.

Euler-Lagrange Equations arise as a necessary condition for a particular curve to be a maximum (or minimum) for certain real-valued