Applied Mathematical Sciences
Volume 90

For further volumes:
http://www.springer.com/series/34

Kenneth R. Meyer
Glen R. Hall
Dan Offin

Introduction to Hamiltonian Dynamical Systems and the N-Body Problem

Second edition

Springer

Kenneth R. Meyer
Department of Mathematics
University of Cincinnati
Cincinnati, OH 45221-0025
USA

Glen R. Hall
Department of Mathematics and Statistics
Boston University
Boston, MA 02215
USA

Dan Offin
Department of Mathematics and Statistics
Queen's University
Kingston, Ontario
Canada

ISBN 978-0-387-09723-7 e-ISBN 978-0-387-09724-4
DOI 10.1007/978-0-387-09724-4

Library of Congress Control Number: 2008940669

Mathematics Subject Classification (2000): 37N05, 70F15, 70Hxx

Printed on acid-free paper

springer.com

Preface to the Second Edition

This new edition expands on some old material and introduces some new subjects. The expanded topics include: parametric stability, logarithms of symplectic matrices, normal forms for Hamiltonian matrices, spacial Delaunay elements, pulsating coordinates, Lyapunov–Chetaev stability applications and more. There is a new section on the Maslov index and a new chapter on variational arguments as applied to the celebrated figure-eight orbit of the 3-body problem.

Still the beginning chapters can serve as a first graduate level course on Hamiltonian dynamical systems, but there is far too much material for a single course. Instructors will have to select chapters to meet their interests and the needs of their class. It will also serve as a reference text and introduction to the literature.

The authors wish to thank their wives and families for giving them the time to work on this project. They acknowledge the support of their universities and various funding agencies including the National Science Foundation, the Taft Foundation, the Sloan Foundation, and the Natural Sciences and Engineering Research Council through the Discovery Grants Program.

This second edition in manuscript form was read by many individuals who made many valuable suggestions and corrections. Our thanks go to Hildeberto Cabral, Scott Dumas, Vadin Fitton, Clarissa Howison, Jesús Palacián, Dieter Schmidt, Jaume Soler, Qiudong Wang, and Patricia Yanguas.

Nonetheless, it is the readers responsibility to inform us of additional errors. Look for email addresses and an errata on MATH.UC.EDU/~MEYER/.

Kenneth R. Meyer
Glen R. Hall
Daniel Offin

Preface to the First Edition

The theory of Hamiltonian systems is a vast subject that can be studied from many different viewpoints. This book develops the basic theory of Hamiltonian differential equations from a dynamical systems point of view. That is, the solutions of the differential equations are thought of as curves in a phase space and it is the geometry of these curves that is the important object of study. The analytic underpinnings of the subject are developed in detail. The last chapter on twist maps has a more geometric flavor. It was written by Glen R. Hall. The main example developed in the text is the classical N-body problem; i.e., the Hamiltonian system of differential equations that describes the motion of N point masses moving under the influence of their mutual gravitational attraction. Many of the general concepts are applied to this example. But this is not a book about the N-body problem for its own sake. The N-body problem is a subject in its own right that would require a sizable volume of its own. Very few of the special results that only apply to the N-body problem are given.

This book is intended for a first course at the graduate level. It assumes a basic knowledge of linear algebra, advanced calculus, and differential equations, but does not assume knowledge of advanced topics such as Lebesgue integration, Banach spaces, or Lie algebras. Some theorems that require long technical proofs are stated without proof, but only on rare occasions. The first draft of the book was written in conjunction with a seminar that was attended by engineering graduate students. The interest and background of these students influenced what was included and excluded.

This book was read by many individuals who made valuable suggestions and many corrections. The first draft was read and corrected by Ricardo Moena, Alan Segerman, Charles Walker, Zhangyong Wan, and Qiudong Wang while they were students in a seminar on Hamiltonian systems. Gregg Buck, Konstantin Mischaikow, and Dieter Schmidt made several suggestions for improvements to early versions of the manuscript. Dieter Schmidt wrote the section on the linearization of the equation of the restricted problem at the five libration points. Robin Vandivier found copious grammatical errors by carefully reading the whole manuscript. Robin deserves a special thanks. We hope that these readers absolve us of any responsibility.

The authors were supported by grants from the National Science Foundation, Defense Advanced Research Project Agency administered by the National Institute of Standards and Technology, the Taft Foundation, and the Sloan Foundation. Both authors were visitors at the Institute for Mathematics and its Applications and the Institute for Dynamics.

Kenneth R. Meyer
Glen R. Hall

Contents

1. Hamiltonian Systems

This chapter defines a Hamiltonian system of ordinary differential equations, gives some basic results about such systems, and presents several classical examples. This discussion is informal. Some of the concepts introduced in the setting of these examples are fully developed later. First, we set forth basic notation and review some basic facts about the solutions of differential equations.

1.1 Notation

$\mathbb{R}$ denotes the field of real numbers, $\mathbb{C}$ the complex field, and $\mathbb{F}$ either $\mathbb{R}$ or $\mathbb{C}$. $\mathbb{F}^n$ denotes the space of all n-dimensional vectors, and, unless otherwise stated, all vectors are column vectors. However, vectors are written as row vectors within the body of the text for typographical reasons. $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ denotes the set of all linear transformations from $\mathbb{F}^n$ to $\mathbb{F}^m$, which are sometimes identified with the set of all $m \times n$ matrices.

Functions are real and smooth unless otherwise stated; smooth means C^∞ or real analytic. If $f(x)$ is a smooth function from an open set in $\mathbb{R}^n$ into $\mathbb{R}^m$, then $\partial f/\partial x$ denotes the $m \times n$ Jacobian matrix

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ & \cdots & \\ & \cdots & \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

If A is a matrix, then A^T denotes its transpose, A^{-1} its inverse, and A^{-T} the inverse transpose. If $f : \mathbb{R}^n \to \mathbb{R}^1$, then $\partial f/\partial x$ is a row vector; let ∇f or $\nabla_x f$ or f_x denote the column vector $(\partial f/\partial x)^T$. Df denotes the derivative of f thought of as a map from an open set in $\mathbb{R}$ into $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. The variable t denotes a real scalar variable called time, and the symbol $\dot{}$ is used for d/dt.

K.R. Meyer et al., *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*, Applied Mathematical Sciences 90, DOI 10.1007/978-0-387-09724-4_1,

1.2 Hamilton's Equations

Newton's second law gives rise to systems of second-order differential equations in $\mathbb{R}^n$ and so to a system of first-order equations in $\mathbb{R}^{2n}$, an even-dimensional space. If the forces are derived from a potential function, the equations of motion of the mechanical system have many special properties, most of which follow from the fact that the equations of motion can be written as a Hamiltonian system. The Hamiltonian formalism is the natural mathematical structure in which to develop the theory of conservative mechanical systems.

A Hamiltonian system is a system of $2n$ ordinary differential equations of the form

$$\begin{gathered} \dot{q} = H_p, \qquad \dot{p} = -H_q, \\ \dot{q}_i = \frac{\partial H}{\partial p_i}(t,q,p), \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}(t,q,p), \quad i = 1,\dots,n, \end{gathered} \tag{1.1}$$

where $H = H(t,q,p)$, called the Hamiltonian, is a smooth real-valued function defined for $(t,q,p) \in \mathcal{O}$, an open set in $\mathbb{R}^1 \times \mathbb{R}^n \times \mathbb{R}^n$. The vectors $q = (q_1,\dots,q_n)$ and $p = (p_1,\dots,p_n)$ are traditionally called the position and momentum vectors, respectively, and t is called time, because that is what these variables represent in the classical examples. The variables q and p are said to be conjugate variables: p is conjugate to q. The concept of conjugate variable grows in importance as the theory develops. The integer n is the number of degrees of freedom of the system.

For the general discussion, introduce the $2n$ vector z, the $2n \times 2n$ skew symmetric matrix J, and the gradient by

$$z = \begin{bmatrix} q \\ p \end{bmatrix}, \quad J = J_n = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}, \quad \nabla H = \begin{bmatrix} \dfrac{\partial H}{\partial z_1} \\ \\ \dfrac{\partial H}{\partial z_{2n}} \end{bmatrix},$$

where 0 is the $n \times n$ zero matrix and I is the $n \times n$ identity matrix. The 2×2 case is special, so sometimes J_2 is denoted by K. In this notation (1.1) becomes

$$\dot{z} = J\nabla H(t,z). \tag{1.2}$$

One of the basic results from the general theory of ordinary differential equations is the existence and uniqueness theorem. This theorem states that for each $(t_0,z_0) \in \mathcal{O}$, there is a unique solution $z = \phi(t,t_0,z_0)$ of (1.2) defined for t near t_0 that satisfies the initial condition $\phi(t_0,t_0,z_0) = z_0$. ϕ is defined on an open neighborhood $\mathcal{Q}$ of $(t_0,t_0,z_0) \in \mathbb{R}^{2n+2}$ into $\mathbb{R}^{2n}$. The function $\phi(t,t_0,z_0)$ is smooth in all its displayed arguments, and so ϕ is C^∞ if the

equations are C^∞, and it is analytic if the equations are analytic. $\phi(t, t_0, z_0)$ is called the general solution. See Chicone (1999), Hubbard and West (1990), or Hale (1972) for details of the theory of ordinary differential equations.

In the special case when H is independent of t, so that $H : \mathcal{O} \to \mathbb{R}^1$ where $\mathcal{O}$ is some open set in $\mathbb{R}^{2n}$, the differential equations (1.2) are autonomous, and the Hamiltonian system is called conservative. It follows that $\phi(t - t_0, 0, z_0) = \phi(t, t_0, z_0)$ holds, because both sides satisfy Equation (1.2) and the same initial conditions. Usually the t_0 dependence is dropped and only $\phi(t, z_0)$ is considered, where $\phi(t, z_0)$ is the solution of (1.2) satisfying $\phi(0, z_0) = z_0$. The solutions are pictured as parameterized curves in $\mathcal{O} \subset \mathbb{R}^{2n}$, and the set $\mathcal{O}$ is called the phase space. By the existence and uniqueness theorem, there is a unique curve through each point in $\mathcal{O}$; and by the uniqueness theorem, two such solution curves cannot cross in $\mathcal{O}$.

An integral for (1.2) is a smooth function $F : \mathcal{O} \to \mathbb{R}^1$ which is constant along the solutions of (1.2); i.e., $F(\phi(t, z_0)) = F(z_0)$ is constant. The classical conserved quantities of energy, momentum, etc. are integrals. The level surfaces $F^{-1}(c) \subset \mathbb{R}^{2n}$, where c is a constant, are invariant sets; i.e., they are sets such that if a solution starts in the set, it remains in the set. In general, the level sets are manifolds of dimension $2n - 1$, and so with an integral F, the solutions lie on the set $F^{-1}(c)$, which is of dimension $2n - 1$. If you were so lucky as to find $2n - 1$ independent integrals, $F_1, \ldots, F_{2n-1}$, then holding all these integrals fixed would define a curve in $\mathbb{R}^{2n}$, the solution curve. In the classical sense, the problem has been integrated.

1.3 The Poisson Bracket

Many of the special properties of Hamiltonian systems are formulated in terms of the Poisson bracket operator, so this operator plays a central role in the theory developed here. Let H, F, and G be smooth functions from $\mathcal{O} \subset \mathbb{R}^1 \times \mathbb{R}^n \times \mathbb{R}^n$ into $\mathbb{R}^1$, and define the Poisson bracket of F and G by

$$\begin{aligned} \{F, G\} &= \nabla F^T J \nabla G = \frac{\partial F}{\partial q}^T \frac{\partial G}{\partial p} - \frac{\partial F}{\partial p}^T \frac{\partial G}{\partial q} \\ &= \sum_{i=1}^{n} \left(\frac{\partial F}{\partial q_i}(t, p, q) \frac{\partial G}{\partial p_i}(t, q, p) - \frac{\partial F}{\partial p_i}(t, q, p) \frac{\partial G}{\partial q_i}(t, q, p) \right). \end{aligned} \tag{1.3}$$

Clearly $\{F, G\}$ is a smooth map from $\mathcal{O}$ to $\mathbb{R}^1$ as well, and one can easily verify that $\{\cdot, \cdot\}$ is skew-symmetric and bilinear. A little tedious calculation verifies Jacobi's identity:

$$\{F, \{G, H\}\} + \{G, \{H, F\}\} + \{H, \{F, G\}\} = 0. \tag{1.4}$$

By a common abuse of notation, let $F(t) = F(t, \phi(t, t_0, z_0))$, where ϕ is the solution of (1.2) as above. By the chain rule,

$$\frac{d}{dt}F(t) = \frac{\partial F}{\partial t}(t, \phi(t, t_0, z_0)) + \{F, H\}(t, \phi(t, t_0, z_0)). \tag{1.5}$$

Hence $dH/dt = \partial H/\partial t$.

Theorem 1.3.1. *Let F, G, and H be as above and independent of time t. Then*

1. *F is an integral for* (1.2) *if and only if $\{F, H\} = 0$.*
2. *H is an integral for* (1.2).
3. *If F and G are integrals for* (1.2)*, then so is $\{F, G\}$.*
4. *$\{F, H\}$ is the time rate of change of F along the solutions of* (1.2).

Proof. (1) follows directly from the definition of an integral and from (1.5). (2) follows from (i) and from the fact that the Poisson bracket is skew-symmetric, so $\{H, H\} = 0$. (3) follows from the Jacobi identity (1.4). (4) follows from (1.5).

In many of the examples given below, the Hamiltonian H is the total energy of a physical system; when it is, the theorem says that energy is a conserved quantity.

In the conservative case when H is independent of t, a critical point of H as a function (i.e., a point where the gradient of H is zero) is an equilibrium (or critical, rest, stationary) point of the system of differential equations (1.1) or (1.2), i.e., a constant solution.

For the rest of this section, let H be independent of t. An equilibrium point ζ of system (1.2) is stable if for every $\epsilon > 0$, there is a $\delta > 0$ such that $\|\zeta - \phi(t, z_0)\| < \epsilon$ for all t whenever $\|\zeta - z_0\| < \delta$. Note that "all t" means both positive and negative t, and that stability is for both the future and the past.

Theorem 1.3.2 (Dirichlet). *If ζ is a strict local minimum or maximum of H, then ζ is stable.*

Proof. Without loss of generality, assume that $\zeta = 0$ and $H(0) = 0$. Because $H(0) = 0$ and 0 is a strict minimum for H, there is an $\eta > 0$ such that $H(z)$ is positive for $0 < \|z\| \leq \eta$. (In the classical literature, one says that H is positive definite.) Let $\kappa = \min(\epsilon, \eta)$ and $M = \min\{H(z) : \|z\| = \kappa\}$, so $M > 0$. Because $H(0) = 0$ and H is continuous, there is a $\delta > 0$ such that $H(z) < M$ for $\|z\| < \delta$. If $\|z_0\| < \delta$, then $H(z_0) = H(\phi(t, z_0)) < M$ for all t. $\|\phi(t, z_0)\| < \kappa \leq \epsilon$ for all t, because if not, there is a time t' when $\|\phi(t', z_0)\| = \kappa$, and $H(\phi(t', z_0)) \geq M$, a contradiction.

1.4 The Harmonic Oscillator

The harmonic oscillator is the second-order, linear, autonomous, ordinary differential equation

$$\ddot{x} + \omega^2 x = 0, \tag{1.6}$$

where ω is a positive constant. It can be written as a system of two first order equations by introducing the conjugate variable $u = \dot{x}/\omega$ and as a Hamiltonian system by letting $H = (\omega/2)(x^2 + u^2)$ (energy in physical problems). The equations become

$$\begin{aligned} \dot{x} &= \ \omega u \ = \frac{\partial H}{\partial u}, \\ \dot{u} &= -\omega x = -\frac{\partial H}{\partial x}. \end{aligned} \tag{1.7}$$

The variable u is a scaled velocity, and thus the x, u plane is essentially the position-velocity plane, or the phase space of physics. The basic existence and uniqueness theorem of differential equations asserts that through each point (x_0, u_0) in the plane, there is a unique solution passing through this point at any particular epoch t_0. The general solutions are given by the formula

$$\begin{bmatrix} x(t, t_0, x_0, u_0) \\ \\ u(t, t_0, x_0, u_0) \end{bmatrix} = \begin{bmatrix} \cos\omega(t - t_0) & -\sin\omega(t - t_0) \\ \\ \sin\omega(t - t_0) & \cos\omega(t - t_0) \end{bmatrix} \begin{bmatrix} x_0 \\ u_0 \end{bmatrix}. \tag{1.8}$$

The solution curves are parameterized circles. The reason that one introduces the scaled velocity instead of using the velocity itself, as is usually done, is so that the solution curves become circles instead of ellipses. In dynamical systems the geometry of this family of curves in the plane is of prime importance. Because the system is independent of time, it admits H as an integral by Theorem 1.3.1 (or note $\dot{H} = \omega x\dot{x} + \omega u\dot{u} = 0$). Because a solution lies in the set where $H =$ constant, which is a circle in the x, u plane, the integral alone gives the geometry of the solution curves in the plane. See Figure 1.1. The origin is a local minimum for H and is stable.

Introduce polar coordinates, $r^2 = x^2 + u^2, \theta = \tan^{-1} u/x$, so that equations (1.7) become

$$\dot{r} = 0, \qquad \dot{\theta} = -\omega. \tag{1.9}$$

This shows again that the solutions lie on circles about the origin because, $\dot{r} = 0$. The circles are swept out with constant angular velocity.

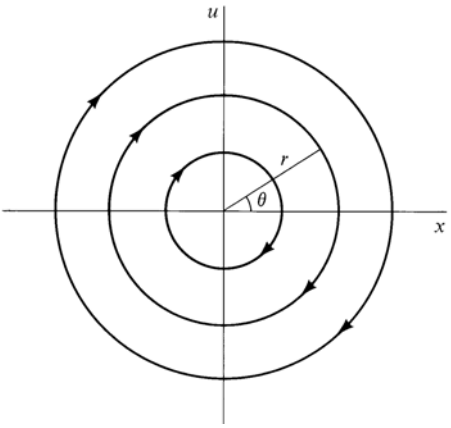


Figure 1.1. Phase portrait of the harmonic oscillator.

1.5 The Forced Nonlinear Oscillator

Consider the system

$$\ddot{x} + f(x) = g(t), \tag{1.10}$$

where x is a scalar and f and g are smooth real-valued functions of a scalar variable. A mechanical system that gives rise to this equation is a spring-mass system. Here, x is the displacement of a particle of mass 1. The particle is connected to a nonlinear spring with restoring force $-f(x)$ and is subject to an external force $g(t)$. One assumes that these are the only forces acting on the particle and, in particular, that there are no velocity-dependent forces acting such as a frictional force.

An electrical system that gives rise to this equation is an LC circuit with an external voltage source. In this case, x represents the charge on a nonlinear capacitor in a series circuit that contains a linear inductor and an external electromotive force $g(t)$. In this problem, assume that there is no resistance in the circuit, and so there are no terms in $\dot{x}$.

This equation is equivalent to the system

$$\dot{x} = y = \frac{\partial H}{\partial y}, \qquad \dot{y} = -f(x) + g(t) = -\frac{\partial H}{\partial x}, \tag{1.11}$$

where

$$H = \frac{1}{2}y^2 + F(x) - xg(t), \qquad F(x) = \int_0^x f(s)ds. \tag{1.12}$$

Many named equations are of this form, for example: (i) the harmonic oscillator: $\ddot{x} + \omega^2 x = 0$; (ii) the pendulum equation: $\ddot{\theta} + \sin\theta = 0$; (iii) the forced Duffing's equation: $\ddot{x} + x + \alpha x^3 = \cos\omega t$.

In the case when the forcing term g is absent, $g \equiv 0$, H is an integral, and the solutions lie in the level curves of H. Therefore, the phase portrait is easily obtained by plotting the level curves. In fact, these equations are integrable in the classical sense that they can be solved "up to a quadrature;" i.e., they are completely solved after one integration or quadrature. Let $h = H(x_0, y_0)$. Solve $H = h$ for y and separate the variables to obtain

$$\begin{aligned} y &= \frac{dx}{dt} = \pm\sqrt{2h - 2F(x)}, \\ t - t_0 &= \pm \int_{x_0}^{x} \frac{d\tau}{\sqrt{2h - 2F(\tau)}}. \end{aligned} \tag{1.13}$$

Thus, the solution is obtained by performing the integration in (1.13) and then taking the inverse of the function so obtained. In general this is quite difficult, but when f is linear, the integral in (1.13) is elementary, and when f is quadratic or cubic, then the integral in (1.13) is elliptic.

1.6 The Elliptic Sine Function

The next example is an interesting classical example. In an effort to extend the table of integrable functions, the elliptic functions were introduced in the nineteenth century. Usually the properties of these functions are developed in advanced texts on complex analysis, but much of the basic properties follow from the elementary ideas in differential equations. Here one example is presented.

Let k be a constant $0 < k < 1$ and $\mathrm{sn}\,(t, k)$ the solution of

$$\ddot{x} + (1 + k^2)x - 2k^2x^3 = 0, \quad x(0) = 0, \quad \dot{x}(0) = 1. \tag{1.14}$$

The function $\mathrm{sn}\,(t, k)$ is called the Jacobi elliptic sine function. Let $y = \dot{x}$. The Hamiltonian, or integral, is

$$2H = y^2 + (1 + k^2)x^2 - k^2x^4 \tag{1.15}$$

and on the solution curve $\mathrm{sn}\,(t, k)$, $2H = 1$, so

$$\dot{\mathrm{sn}}^2 = (1 - \mathrm{sn}^2)(1 - k^2\mathrm{sn}^2). \tag{1.16}$$

The phase portrait of (1.14) is the level line plot of H. To find this plot, first graph

$$\ell(x) = 2h - (1 + k^2)x^2 + k^2x^4 = (2h - 1) + (1 - x^2)(1 - k^2x^2).$$

Then take square roots by plotting $y^2 = \ell(x)$ to obtain the phase portrait of (1.14) as shown in Figure 1.2.

The solution curve $\mathrm{sn}\,(t,k)$ lies in the connected component of $2H = 1$ which contains $x = 0$, $y = \dot{x} = 1$, i.e., the closed curve encircling the origin illustrated by the darker oval in Figure 1.2. The solution $\mathrm{sn}\,(t,k)$ lies on a closed level line that does not contain an equilibrium point, therefore it must be a periodic function.

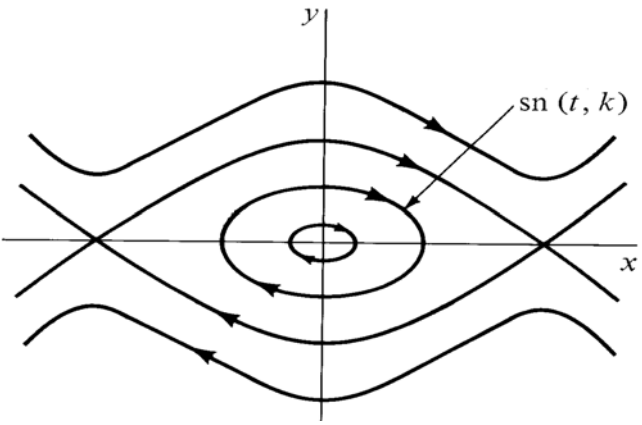


Figure 1.2. Phase portrait of the elliptic sine function.

Both $\mathrm{sn}\,(t,k)$ and $-\mathrm{sn}\,(-t,k)$ satisfy (1.14), and so by the uniqueness theorem for ordinary differential equations, $\mathrm{sn}\,(t,k) = -\mathrm{sn}\,(-t,k)$, i.e., sn is odd in t. The curve defined by sn goes through the points $x = \pm 1, y = 0$ also. As t increases from zero, $\mathrm{sn}\,(t,k)$ increases from zero until it reaches its maximum value of 1 after some time, say a time κ . (Classically, the constant κ is denoted by K.) Because $\mathrm{sn}\,(\pm\kappa,k) = \pm 1$ and $\dot{\mathrm{sn}}\,(\pm\kappa,k) = 0$ and both $\mathrm{sn}\,(t+\kappa,k)$ and $-\mathrm{sn}\,(t-\kappa,k)$ satisfy the equation in (1.14), by uniqueness of the solutions of differential equations it follows that $\mathrm{sn}\,(t+\kappa,k) = -\mathrm{sn}\,(t-\kappa,k)$, or that sn is 4κ periodic and odd harmonic in t. Thus the Fourier series expansion of sn only contains terms in $\sin(j2\pi t/4\kappa)$ where j is an odd integer.

It is clear that sn is increasing for $-\kappa < t < \kappa$. Equation (1.14) implies $\ddot{\mathrm{sn}} > 0$ (so sn is convex) for $-\kappa < t < 0$, and it also implies $\ddot{\mathrm{sn}} < 0$ (so sn is concave) for $0 < t < \kappa$. Thus, sn has the same basic symmetry properties as the sine function. It is also clear from the equations that $\mathrm{sn}\,(t,k) \to \sin t$ and $\kappa \to \pi/2$ as $k \to 0$. The graph of sn, (t,k) has the same general form as $\sin t$ with 4κ playing the role of 2π.

The function $\kappa(k)$ is investigated in the problems. Classical handbooks contain tables of values of the sn function, and computer algebra systems such as Maple have these functions. Thus one knows almost as much about $\mathrm{sn}\,(t,k)$

as about $\sin t$. Your list of elementary functions should contain $\mathrm{sn}\,(t,k)$. In the problems, you are asked to solve the pendulum equation with your new elementary function.

There are two other Jacobi elliptic functions that satisfy equations similar to (1.14). They were introduced in order to extend the number of functions that can be integrated. In fact, with the three Jacobi elliptic functions, all equations of the form (1.10) with $g = 0$ and $f(x)$ a quadratic or cubic polynomial can be solved explicitly. A different and slightly more detailed discussion is found in Meyer (2001), and the classic text Modern Analysis by Whittaker and Watson (1927) has a complete discussion of the Jacobi elliptic functions. Many of the formulas will remind one of trigonometry.

1.7 General Newtonian System

The n-dimensional analog of (1.10) is

$$M\ddot{x} + \nabla F(x) = g(t), \tag{1.17}$$

where x is an n-vector, M is a nonsingular, symmetric $n \times n$ matrix, F is a smooth function defined on an open domain $\mathcal{O}$ in $\mathbb{R}^n$, ∇F is the gradient of F, and g is a smooth n-vector valued function of t, for t in some open set in $\mathbb{R}^1$. Let $y = M\dot{x}$. Then (1.17) is equivalent to the Hamiltonian system

$$\dot{x} = \frac{\partial H}{\partial y} = M^{-1}y, \qquad \dot{y} = -\frac{\partial H}{\partial x} = -\nabla F(x) + g(t), \tag{1.18}$$

where the Hamiltonian is

$$H = \frac{1}{2}y^T M^{-1} y + F(x) - x^T g(t). \tag{1.19}$$

If x represents the displacement of a particle of mass m, then $M = mI$ where I is the identity matrix, y is the linear momentum of the particle, $\frac{1}{2}y^T M^{-1} y$ is the kinetic energy, $g(t)$ is an external force, and F is the potential energy. If $g(t) \equiv 0$, then H is an integral and is total energy. This terminology is used in reference to nonmechanical systems of the form (1.17) also. In order to write (1.18) as a Hamiltonian system, the correct choice of the variable conjugate to x is $y = M\dot{x}$, the linear momentum, and not $\dot{x}$, the velocity.

In the special case when $g \equiv 0$, a critical point of the potential is a critical point of H and hence an equilibrium point of the Hamiltonian system of equations (1.18). In many physical examples, M is positive definite. In this case, if x' is a local minimum for the potential F, then $(x', 0)$ is a local minimum for H and therefore a stable equilibrium point by Theorem 1.3.2.

It is tempting to think that if x' is a critical point of F and not a minimum of the potential, then the point $(x', 0)$ is an unstable equilibrium point. This is not true. See Laloy (1976) and Chapter 13 for a discussion of stability questions for Hamiltonian systems.

1.8 A Pair of Harmonic Oscillators

Consider a pair of harmonic oscillators

$$\ddot{x} + \omega^2 x = 0, \qquad \ddot{y} + \mu^2 y = 0,$$

which as a system becomes the Hamiltonian system

$$\begin{aligned} \dot{x} &= \omega u = \frac{\partial H}{\partial u}, & \dot{y} &= \mu v = \frac{\partial H}{\partial v}, \\ \dot{u} &= -\omega x = -\frac{\partial H}{\partial x}, & \dot{v} &= -\mu y = -\frac{\partial H}{\partial y}, \end{aligned} \tag{1.20}$$

where the Hamiltonian is

$$H = \frac{\omega}{2}(x^2 + u^2) + \frac{\mu}{2}(y^2 + v^2).$$

In polar coordinates

$$r^2 = \frac{\omega}{2}(x^2 + u^2), \qquad \theta = \tan^{-1} u/x,$$

$$\rho^2 = \frac{\mu}{2}(y^2 + v^2), \qquad \phi = \tan^{-1} v/y,$$

the equations become

$$\begin{aligned} \dot{r} &= 0, & \dot{\theta} &= -\omega, \\ \dot{\rho} &= 0, & \dot{\phi} &= -\mu, \end{aligned} \tag{1.21}$$

and they admit the two integrals

$$I_1 = r^2 = (\omega/2)(x^2 + u^2), \qquad I_2 = \rho^2 = (\mu/2)(y^2 + v^2). \tag{1.22}$$

In many physical problems, these equations are only the first approximation. The full system does not admit the two individual integrals (energies), but does admit H as an integral which is the sum of the individual integrals. Think, for example, of a pea rolling around in a bowl; the linearized system at the minimum would be of the form (1.20). In this case, $H^{-1}(1)$ is an invariant set for the flow, which is an ellipsoid and topologically a 3-sphere.

Consider the general solution through $r_0, \theta_0, \rho_0, \phi_0$ at epoch $t = 0$. The solutions with $r_0 = 0$ and $\rho_0 > 0$ or $\rho_0 = 0$ and $r_0 > 0$ lie on circles and correspond to periodic solutions of period $2\pi/\mu$ and $2\pi/\omega$, respectively. These periodic solutions are special and are usually called the normal modes.

The set where $r = r_0 > 0$ and $\rho = \rho_0 > 0$ is an invariant torus for (1.20) or (1.21). Angular coordinates on this torus are θ and ϕ, and the equations are

$$\dot{\theta} = -\omega, \qquad \dot{\phi} = -\mu, \tag{1.23}$$

the standard linear equations on a torus. See Figure 1.3.

If ω/μ is rational, then $\omega = p\tau$ and $\mu = q\tau$, where p and q are relatively prime integers and τ is a nonzero real number. In this case the solution of (1.23) through θ_0, ϕ_0 at epoch $t = 0$ is $\theta(t) = \theta_0 - \omega t$, $\phi(t) = \phi_0 - \mu t$, and so if $T = 2\pi/\tau$, then $\theta(T) = \theta_0 + p2\pi$ and $\phi(T) = \phi_0 + q2\pi$. That is, the solution is periodic with period T on the torus, and this corresponds to periodic solutions of (1.20).

If ω/μ is irrational, then none of the solutions is periodic. In fact, the solutions of (1.23) are dense lines on the torus see Section 1.9), and this corresponds to the fact that the solutions of (1.20) are quasiperiodic but not periodic.

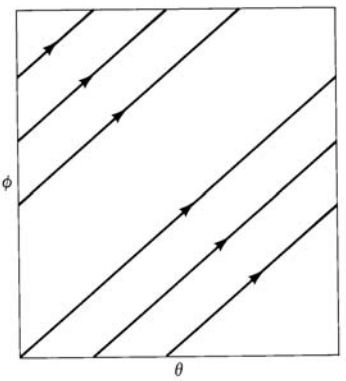


Figure 1.3. Linear flow on the torus.

We can use polar coordinates to introduce coordinates on the sphere, provided we are careful to observe the conventions of polar coordinates: (i) $r \geq 0$, (ii) θ is defined modulo 2π, and (iii) $r = 0$ corresponds to a point. That is, if we start with the rectilinear strip $r \geq 0$, $0 \leq \theta \leq 2\pi$, then identify the $\theta = 0$ and $\theta = 2\pi$ edges to get a half-closed annulus, and finally if we identify the circle $r = 0$ with a point, then we have a plane (Figure 1.4).

Starting with the polar coordinates r, θ, ρ, ϕ for $\mathbb{R}^4$, we note that on the 3-sphere, $E = r^2 + \rho^2 = 1$, so we can discard ρ and have $0 \leq r \leq 1$. We use r, θ, ϕ as coordinates on S^3. Now r, θ with $0 \leq r \leq 1$ are just polar coordinates for the closed unit disk. For each point of the open disk, there is a circle with coordinate ϕ (defined mod 2π), but when $r = 1, \rho = 0$, so the circle collapses to a point over the boundary of the disk. The geometric model of S^3 is given

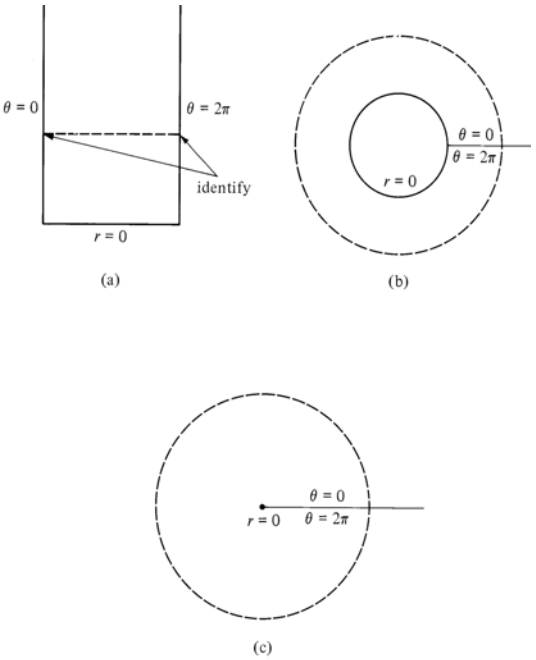
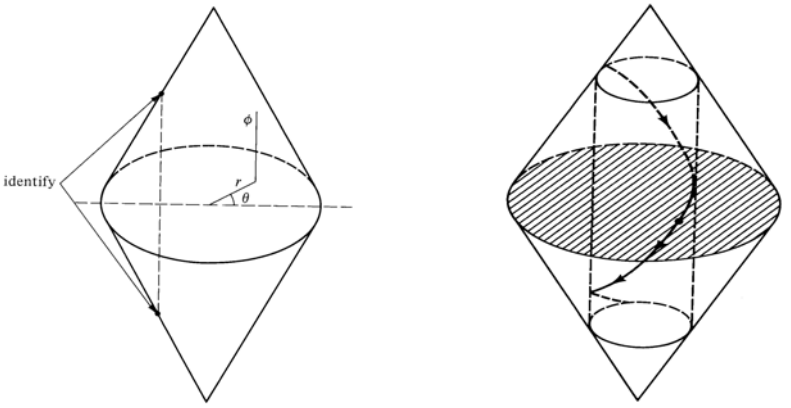


Figure 1.4. The polar coordinate conventions.

by two solid cones with points on the boundary cones identified as shown in Figure 1.5a. Through each point in the open unit disk with coordinates r, θ there is a line segment (the dashed line) perpendicular to the disk. The angular coordinate ϕ is measured on this segment: $\phi = 0$ is the disk, $\phi = \pi$ is the upper boundary cone, and $\phi = -\pi$ is the lower boundary cone. Each point on the upper boundary cone with coordinates $r, \theta, \phi = \pi$ is identified with the point on the lower boundary cone with coordinates $r, \theta, \phi = -\pi$. From this model follows a series of interesting geometric facts.

For α, $0 < \alpha < 1$, the set where $r = \alpha$ is a 2-torus in the 3-sphere, and for $\alpha = 0$ or 1, the set $r = \alpha$ is a circle. Because r is an integral for the pair of oscillators, these tori and circles are invariant sets for the flow defined by the harmonic oscillators. The two circles $r = 0, 1$ are periodic solutions, called the normal modes. The two circles are linked in S^3, i.e., one of the circles intersects a disk bounded by the other circle in an algebraically nontrivial way. The circle where $r = 1$ is the boundary of the shaded disk in Figure 1.5b, and the circle $r = 0$ intersects this disk once. It turns out that the number of intersections is independent of the bounding disk provided one counts the intersections algebraically.

Consider the special case when $\omega = \mu = 1$. In this case every solution is periodic, and so its orbit is a circle in the 3-sphere. Other than the two special



(a) A model of S^3 (b) An orbit on S^3

Figure 1.5. S^3 as a circle bundle over S^2.

circles, on each orbit as θ increases by 2π, so does ϕ. Thus each such orbit hits the open disk where $\phi = 0$ (the shaded disk in Figure 1.5) in one point. We can identify each such orbit with the unique point where it intersects the disk. One special orbit meets the disk at the center, so we can identify it with the center. The other special orbit is the outer boundary circle of the disk which is a single orbit. When we identify this circle with a point, the closed disk whose outer circle is identified with a point becomes a 2-sphere.

Theorem 1.8.1. *The 3-sphere, S^3, is the union of circles. Any two of these circles are linked. The quotient space obtained by identifying a circle with a point is a 2-sphere (the Hopf fibration of S^3).*

Let D be the open disk $\phi = 0$, the shaded disk in Figure 1.5. The union of all the orbits that meet D is a product of a circle and a 2-disk, so each point not on the special circle $r = 1$ lies in an open set that is the product of a 2-disk and a circle. By reversing r and ρ in the discussion given above, the circle where $r = 1$ has a similar neighborhood. So locally the 3-sphere is the product of a disk and a circle, but the sphere is not the product of a 2-manifold and a circle. (The sphere has a trivial fundamental group, but such a product would not.)

When $\omega = p$ and $\mu = q$ with p and q relatively prime integers, all solutions are periodic, and the 3-sphere is again a union of circles, but it is not locally a product near the special circles. The nonspecial circles are p, q-torus knots. They link p times with one special circle and q times with the other.

These links follow by a slight extension of the ideas of the previous proposition. A p, q-torus knot is a closed curve that wraps around the standard torus in $\mathbb{R}^3$ in the longitudinal direction p times and in the meridional direction q times. If p and q are different from 1, the knot is nontrivial.

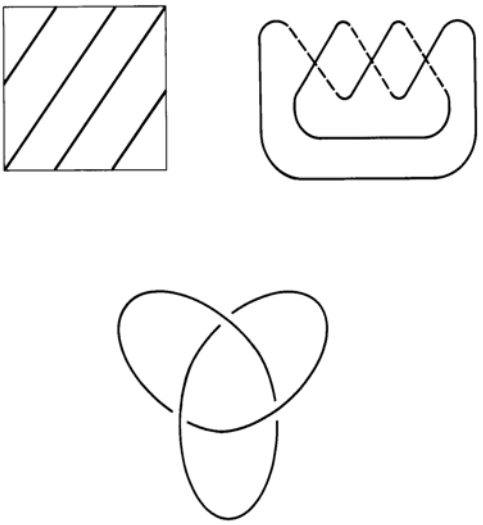


Figure 1.6. The trefoil as a toral knot.

Figure 1.6 shows that the 3,2 torus knot is the classical trefoil or cloverleaf knot. The first diagram in Figure 1.6 is the standard model of a torus: a square with opposite sides identified. The line with slope 3/2 is shown wrapping three times around one way and twice around the other. Think of folding the top half of the square back and around and then gluing the top edge to the bottom to form a cylinder. Add two extra segments of curves to connect the right and left ends of the curve to get the second diagram in Figure 1.6. Smoothly deform this to get the last diagram in Figure 1.6, the standard presentation of the trefoil. See Rolfsen (1976) for more information on knots.

1.9 Linear Flow on the Torus

In order to show that the solutions of (1.23) on the torus are dense when ω/μ is irrational, the following simple lemmas from number theory are needed.

Lemma 1.9.1. *Let δ be any irrational number. Then for every $\epsilon > 0$, there exist integers q and p such that*

$$| q\delta - p | < \epsilon. \tag{1.24}$$

Proof. Case 1: $0 < \delta < 1$. Let $N \geq 2$ be an integer and $S_N = \{s\delta - r : 1 \leq s, r \leq N\}$. For each element of this set we have $| s\delta - r | <$ N. Because δ

is irrational, there are N^2 distinct members in the set S_N; so at least one pair is less than $4/N$ apart. (If not, the total length would be greater than $(N^2-1)4/N > 2N$.) Call this pair $s\delta - r$ and $s'\delta - r'$. Thus

$$0 <\mid (s-s')\delta - (r-r') \mid < \frac{4}{N} < \frac{4}{\mid s-s' \mid}. \tag{1.25}$$

Take $N > 4/\epsilon, q = s - s'$ and $p = r' - r$ to finish this case. The other cases follow from the above. If $-1 < \delta < 0$, then apply the above to $-\delta$; and if $\mid \delta \mid > 1$, apply the above to $1/\delta$.

Lemma 1.9.2. *Let δ be any irrational number and ξ any real number. Then for every $\epsilon > 0$ there exist integers p and q such that*

$$\mid q\delta - p - \xi \mid < \epsilon. \tag{1.26}$$

Proof. Let p' and q' be as given in Lemma 1.9.1, so $\eta = q'\delta - p'$ satisfies $0 <\mid \eta \mid < \epsilon$. There is a integer m such that $\mid m\eta - \xi \mid < \epsilon$. The lemma follows by taking $q = mq'$ and $p = mp'$.

Theorem 1.9.1. *Let ω/μ be irrational. Then the solution curves defined by Equations* (1.23) *are dense on the torus.*

Proof. Measure the angles in revolutions instead of radians so that the angles θ and ϕ are defined modulo 1 instead of 2π. The solution of equations (1.23) through $\theta = \phi = 0$ at $t = 0$ is $\theta(t) = \omega t, \phi(t) = \mu t$. Let $\epsilon > 0$ and ξ be given. Then $\theta \equiv \xi$ and $\phi \equiv 0 \bmod 1$ is an arbitrary point on the circle $\phi \equiv 0 \bmod 1$ on the torus. Let $\delta = \omega/\mu$ and p, q be as given in Lemma 2. Let $\tau = q/\mu$, so $\theta(\tau) = \delta q, \phi(\tau) = q$. Thus, $\mid \theta(\tau) - p - \xi \mid < \epsilon$, but because p is an integer, this means that $\theta(\tau)$ is within ϵ of ξ; so the solution through the origin is dense on the circle $\phi \equiv 0 \bmod 1$. The remainder of the proof follows by translation.

1.10 Euler–Lagrange Equations

Many of the laws of physics can be given as minimizing principles and this led the theologian-mathematician Leibniz to say that we live in the best of all possible worlds. In more modern times and circumstances, the physicist Richard Feynman once quoted that of all mathematical-physical principles, the principle of least action is one that he has pondered most frequently.

Under mild smoothness conditions, one shows in the calculus of variations that minimizing the curve functional with fixed boundary constraints

$$F(q) = \int_\alpha^\beta L(q(t), \dot{q}(t))\, dt, \qquad q(\alpha) = q_\alpha, \qquad q(\beta) = q_\beta$$

leads to a function $q : [\alpha, \beta] \to \mathbb{R}^n$ satisfying the Euler–Lagrange equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = 0. \tag{1.27}$$

These equations are also known as Euler's equations.

Here we use the symbol $\dot{q}$ with two meanings. The function L is a function of two variables and these two variables are denoted by $q, \dot{q}$, so $\partial L/\partial \dot{q}$ denotes the partial derivative of L with respect to its second variable. A solution of (1.27) is a smooth function of t, denoted by $q(t)$, whose derivative with respect to t is $\dot{q}(t)$.

In particular, if $q, \dot{q}$ are the position-velocity of a mechanical system subject to a system of holonomic constraints and $K(\dot{q})$ is its kinetic energy, $P(q)$ its potential energy, and $L = K - P$ the Lagrangian then (1.27) is the equation of motion of the system — see Arnold (1978), Siegel and Moser (1971), or almost any advanced texts on mechanics.

More generally, any critical point of the action functional $F(\cdot)$ leads to the same conclusion concerning the critical function $q(\cdot)$. Moreover, the boundary conditions for the variational problem may be much more general, including the case of periodic boundary conditions, which would replace the fixed endpoint condition with the restriction on the class of functions

$$q(\alpha) = q(\beta)$$

This is an important generalization, in as much as all the periodic solutions of the N-body problem can be realized as critical points of the action, subject to the periodic boundary condition. In fact, this observation leads one to look for such periodic solutions directly by finding appropriate critical points of the action functional, rather than by solving the boundary value problem connected with the Euler equations. This is called the direct method of the calculus of variations, which is a global method in that it does not require nearby known solutions for its application. This method has recently helped the discovery of some spectacular new periodic solutions of the N-body problem that are far from any integrable cases and which are discussed in subsequent chapters. We give a very simple example of this method below, together with some extensions of this method to the question of global stability of periodic solutions.

Here are the ingredients of the argument that relates the critical points of F to the Euler–Lagrange equations. Suppose that q_ϵ is a one parameter curve of functions through the critical function q that satisfies the boundary constraints. That is, $q_0(t) = q(t), \quad \alpha \le t \le \beta$, and $q_\epsilon(\alpha) = q_\alpha, \quad q_\epsilon(\beta) = q_\beta$ in the case of fixed boundary conditions, or $q_\epsilon(\alpha) = q_\epsilon(\beta)$ in the case of periodic conditions. In either of these cases, one would naturally infer that the composite function $g(\epsilon) = F(q_\epsilon)$ has a critical point at $\epsilon = 0$. Assuming that we are able to differentiate under the integral sign, and that the variation vector field

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

is smooth, we find that

$$\begin{aligned} dF(q)\cdot\xi = \left.\frac{\partial}{\partial\epsilon}F(q_\epsilon)\right|_{\epsilon=0} &= \int_\alpha^\beta \left(\frac{\partial L}{\partial x}\cdot\xi + \frac{\partial L}{\partial\dot{x}}\cdot\dot{\xi}\right)dt \\ &= \left.\frac{\partial L}{\partial\dot{x}}\cdot\xi\right|_\alpha^\beta + \int_\alpha^\beta \left(-\frac{d}{dt}\frac{\partial L}{\partial\dot{x}} + \frac{\partial L}{\partial x}\right)\cdot\xi\, dt. \end{aligned} \tag{1.28}$$

The last line of (1.28) is done using an integration by parts. It is not difficult to see that if the function q is critical for the functional F with either set of boundary conditions, then the boundary terms and the integral expression must vanish independently for an arbitrary choice of the variation vector field $\xi(t)$. This leads to two conclusions: first, that the Euler–Lagrange equations (1.27) must vanish identically on the interval $\alpha \le t \le \beta$ and second, that the transversality conditions

$$\left.\frac{\partial L}{\partial\dot{x}}\cdot\xi\right|_\alpha^\beta = 0 \tag{1.29}$$

should also hold for the critical function q at the endpoints α, β. In the case of fixed boundary conditions, these transversality conditions don't give any additional information because $\xi(\alpha) = \xi(\beta) = 0$. In the case of periodic boundary conditions, they imply that

$$\frac{\partial L}{\partial\dot{q}}(\alpha) = \frac{\partial L}{\partial\dot{q}}(\beta), \tag{1.30}$$

because $\xi(\alpha) = \xi(\beta)$. As we show below, this guarantees that a critical point of the action functional with periodic boundary conditions, is just the configuration component of a periodic solution of Hamilton's equations.

We have shown in (1.28) that we can identify critical points of the functional $F(\cdot)$ with solutions of the Euler equations (1.27) subject to various boundary constraints. One powerful and important application of this is that the Euler–Lagrange equations are invariant under general coordinate transformations.

Proposition 1.10.1. *If the transformation $(x, \dot{x}) \to (q, \dot{q})$ is a local diffeomorphism with*

$$q = q(x), \quad \dot{q} = \frac{\partial q}{\partial x}(x)\cdot\dot{x},$$

then the Euler–Lagrange equations (1.27) *transform into an equivalent set of Euler–Lagrange equations*

$$\frac{d}{ds}\frac{\partial\tilde{L}}{\partial\dot{x}} - \frac{\partial\tilde{L}}{\partial x} = 0,$$

where the new Lagrangian is defined by the coordinate transformation

$$\tilde{L}(x, \dot{x}) = L(q(x), \frac{\partial q}{\partial x}(x) \cdot \dot{x}).$$

Proof. The argument rests on two simple observations. First, the condition that $F(q)$ take a critical value is independent of coordinates; and second, the functional $F(q)$ transforms in a straightforward manner

$$\begin{aligned} F(q(t)) &= \int_\alpha^\beta L(q(t), \dot{q}(t))dt \\ &= \int_\alpha^\beta L(q(x(t)), \frac{\partial q}{\partial x}(x(t)) \cdot \dot{x}(t))dt \\ &= \int_\alpha^\beta \tilde{L}(x(t), \dot{x}(t))dt = \tilde{F}(x(t)). \end{aligned}$$

From this, we conclude that the critical points of $F(\cdot)$ correspond to critical points of $\tilde{F}(\cdot)$ under the coordinate transformation. The conclusion of the proposition follows, because we have shown in (1.28) that critical points of $F(\cdot)$ are solutions of the Euler equations for the Lagrangian L, and critical points of $\tilde{F}(\cdot)$ are solutions of the Euler equations for the Lagrangian $\tilde{L}$.

Sometimes L depends on t and we wish to change the time variable also. By the same reasoning, if the transformation $(x, x', s) \to (q, \dot{q}, t)$ is

$$q = q(x, s), \quad t = t(x, s), \quad \dot{q} = \dot{q}(x, x', s) = \frac{q_x(x, s)x' + q_s(x, s)}{t_x(x, s)x' + t_s(x, s)},$$

then the Euler–Lagrange equations (1.27) become

$$\frac{d}{ds}\frac{\partial \tilde{L}}{\partial x'} - \frac{\partial \tilde{L}}{\partial x} = 0,$$

where $' = d/ds$ and

$$\tilde{L}(x, x', s) = L(q(x, s), \dot{q}(x, x', s), t(x, s)).$$

We consider one interesting example here, whereby the variational structure of certain solutions is directly tied to the stability type of these solutions. We follow this thread of an idea in later examples, especially when we apply the variational method to finding symmetric periodic solutions of the N-body problem. The mathematical pendulum is given by specifying a constrained mechanical system in the plane with Cartesian coordinates (x, y). The gravitational potential energy is $U(x, y) = mgy$ and the kinetic energy $K(\dot{x}, \dot{y}) = \frac{1}{2}(\dot{x}^2 + \dot{y}^2)$. The constraint requires the mass m to lie at a fixed distance l from the point $(0, l)$ so that $x^2 + (y - l)^2 = l^2$. Introducing a local angular coordinate θ mod 2π on the circle $x^2 + (y - l)^2 = l^2$ and expressing

the Lagrangian in these coordinates we find the Lagrangian and the resulting Euler–Lagrange equations,

$$L(\theta, \dot{\theta}) = \frac{1}{2} ml^2 \dot{\theta}^2 + mgl(1 + \cos(\theta)), \qquad ml^2 \ddot{\theta} = -mgl \sin(\theta).$$

The equations in θ follow by the invariance of Euler–Lagrange equations (1.27), see Proposition (1.10.1)) The factor mgl is subtracted from the potential to make the action positive, and doesn't affect the resulting differential equations. The action of the variational problem is the integral of the Lagrangian, so we study the periodic problem

$$F(q) = \int_0^T \left(\frac{1}{2} ml^2 \dot{q}^2 + mgl(1 + \cos(q)) \right) dt, \qquad q(0) = q(T).$$

We make the simple observation that the absolute minimizer of the action corresponds to a global maximum of the potential, and the global minimum of the potential corresponds to a mountain pass critical point of the action functional

$$F(\pm\pi) \leq F(q), \qquad F(0) = \min \max_{\deg q=1} F(q).$$

The first inequality may be easily verified, because the kinetic energy is positive and the potential takes a maximum value at $\pm\pi$. In the second case, the maximum is taken with respect to loops in the configuration variable, which make one circuit of the point 0 before closing. This is described by the topological degree = 1. The minimum is then taken over all such loops, including the limit case when the loop is stationary at the origin.

It is interesting to observe here that the global minimum of the action functional corresponds to a hyperbolic critical point, and the stable critical point (see Dirichlet's theorem (1.3.2)) corresponds to a mountain pass type critical point. This fact is not isolated, and we discuss a theory to make this kind of prediction concerning stability and instability in much more general settings when we discuss the Maslov index in Section 4.5.

One could consider the forced pendulum equations, as was done in Section 1.5. Here the analysis and the results become essentially more interesting, because there are no longer any equilibrium solutions; however, the direct method of the calculus of variations leads to some very interesting global results for this simple problem, which we describe briefly. The Euler equation and the action functional become

$$ml^2 \ddot{\theta} = -mgl \sin(\theta) + f(t), \quad f(t+T) = f(t),$$

$$F(q) = \int_0^T \left(\frac{1}{2} ml^2 \dot{q}^2 + mgl(1 + \cos(q)) + qf(t) \right) dt, \quad q(0) = q(T).$$

In this problem, the question of stable and unstable periodic solutions becomes an interesting nonelementary research topic. The first question one