Developments in Mathematics

Russell Johnson
Rafael Obaya
Sylvia Novo
Carmen Núñez
Roberta Fabbri

Nonautonomous Linear Hamiltonian Systems: Oscillation, Spectral Theory and Control

Springer

Developments in Mathematics

VOLUME 36

Series editors

More information about this series at http://www.springer.com/series/5834

Russell Johnson • Rafael Obaya • Sylvia Novo •
Carmen Núñez • Roberta Fabbri

Nonautonomous Linear Hamiltonian Systems: Oscillation, Spectral Theory and Control

Springer

Russell Johnson
Università degli Studi di Firenze
Firenze, Italy

Rafael Obaya
Universidad de Valladolid
Valladolid, Spain

Sylvia Novo
Universidad de Valladolid
Valladolid, Spain

Carmen Núñez
Universidad de Valladolid
Valladolid, Spain

Roberta Fabbri
Università degli Studi di Firenze
Firenze, Italy

ISSN 1389-2177 ISSN 2197-795X (electronic)
Developments in Mathematics
ISBN 978-3-319-29023-2 ISBN 978-3-319-29025-6 (eBook)
DOI 10.1007/978-3-319-29025-6

Library of Congress Control Number: 2016935697

Mathematics Subject Classification: 37B55, 34C10, 54H20, 34D08, 34D09, 34H05, 49N05, 37D25, 53D12, 37C55

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

This book is devoted to a study of the oscillation theory of nonautonomous linear Hamiltonian differential systems and that of a spectral theory which is adapted to such systems. Systematic use will be made of basic facts concerning Lagrange subspaces of $\mathbb{R}^{2n}$ and argument functions on the set of symplectic matrices. We will also consistently apply some fundamental methods of topological dynamics and of ergodic theory, including Lyapunov exponents, exponential dichotomies, and rotation numbers. Further, we will show that our results concerning oscillation theory can be fruitfully applied to several basic issues in the theory of linear-quadratic control systems with time-varying coefficients.

Nonautonomous Oscillation Theory

In due course, we will give an outline of the specific problems, methods, and results to be discussed in the body of the book. Before doing that, it seems appropriate to collocate them in a priori way in the vast and nonhomogeneous area called oscillation theory of ordinary differential equations. In fact, the word "oscillation" has various meanings in this context. For example, it can refer to the study of the zeroes contained in some interval $\mathcal{I} \subseteq \mathbb{R}$ of a solution of an ordinary differential equation (ODE). In the case of a two-dimensional ODE, it can refer to the variation of the polar angle along a solution, i.e., to the "rotation" associated to that solution. Still again, it may indicate one of the many themes encountered in the study of the periodic solutions of an ordinary differential equation.

This book is about "rotation." Let us try to be a bit more precise. We will focus attention on various issues concerning the solutions of a linear Hamiltonian differential system

$$\mathbf{z}' = H(t)\,\mathbf{z}\,, \tag{1}$$

where $\mathbf{z} \in \mathbb{R}^{2n}$ and $t \in \mathbb{R}$. The coefficient $H(\cdot)$ is a bounded measurable real $2n \times 2n$ matrix-valued function satisfying the symplectic condition $(JH)^T(t) = JH(t)$ for all $t \in \mathbb{R}$, where the "T" indicates the transpose and $J = \left[\begin{smallmatrix} 0_n & -I_n \\ I_n & 0_n \end{smallmatrix}\right]$ is the usual $2n \times 2n$ antisymmetric matrix: I_n is the $n \times n$ identity matrix and 0_n the $n \times n$ zero matrix. Generally speaking, we will be interested in the "rotation" of the solutions of (1). Of course, this notion is initially problematic because it is not immediately clear how to define it precisely, especially if $n \geq 2$. One of our main goals will be to do this. It will turn out that our concept of rotation is closely related to a more or less standard notion of a "point of verticality" of a solution of (1), namely, a focal point. It will also turn out that the concept of rotation considered here can be used to study some basic questions in spectral theory, which are formulated in terms of equation (1) and which will be discussed shortly.

Equation (1) is of course very significant. As a special case, one can set $\mathbf{z} = \left[\begin{smallmatrix} \mathbf{x} \\ \mathbf{y} \end{smallmatrix}\right]$ for $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, and

$$H(t) = \begin{bmatrix} 0_n & I_n \\ G(t) & 0_n \end{bmatrix},$$

where $G^T = G$ is a real symmetric $n \times n$ matrix-valued function. Then (1) is equivalent to the second-order system

$$\mathbf{x}'' = G(t)\,\mathbf{x}\,, \tag{2}$$

which is often encountered in the study of mechanical systems near an equilibrium. Another special case is obtained by setting $n = 1$ and

$$H(t) = \begin{bmatrix} 0 & 1/p(t) \\ g(t) - \lambda\, d(t) & 0 \end{bmatrix}$$

for a real parameter λ; in this case (1) is equivalent to the classical Sturm–Liouville problem

$$-\,(p\,x')' + g(t)\,x = \lambda\, d(t)\,x\,. \tag{3}$$

Problem (3) has been studied with success from various points of view for over 150 years. The number and the location of the zeroes of a solution $x(\cdot)$ are a recurring theme. Information concerning these zeroes has implications for the spectral problem obtained by varying λ and by imposing boundary conditions, for example, of Dirichlet type: $x(a) = x(b) = 0$ where $a < b \in \mathbb{R}$. Then, as is well known, if p, g, and d satisfy certain general hypotheses, then the nth eigenfunction of (3) has $n - 1$ zeroes in (a, b), for $n = 1, 2, \ldots$

A more general spectral problem is obtained by using (1) as a point of departure. One introduces a parameter $\lambda \in \mathbb{R}$ and a positive semidefinite real weight function $\Gamma(t)$ in (1), so as to obtain

$$\mathbf{z}' = \big(H(t) + \lambda\, J^{-1}\, \Gamma(t)\big)\, \mathbf{z}\,. \tag{4}$$

This problem was studied systematically by Atkinson in [5]. It is noteworthy that if Γ is semidefinite but not everywhere definite, then the study of the boundary-value problem associated to (4) cannot be naturally carried out using standard functional-analytic techniques (due to the fact that one cannot multiply (4) by Γ^{-1}). However, in [5], one finds an "Atkinson condition" which, when imposed on (4), allows the development of a satisfactory spectral theory for (4).

Another of our goals is to show that our oscillation theory of (1) can be fruitfully applied to the spectral problem (4) especially when "the boundary conditions are imposed at $t = \pm\infty$," i.e., when (4) is considered on the whole line. Let us explain some of the issues involved in relating oscillation theory and spectral theory in the context of problem (4). Consider for a moment the version of (3) obtained by setting $p = d \equiv 1$:

$$-\,x'' + g(t)\,x = \lambda\,x\,. \tag{5}$$

This is the Schrödinger equation with potential $g(t)$ (a most important ordinary differential equation, due to its basic role in one-dimensional quantum mechanics). Fix $\lambda \in \mathbb{R}$, and consider a solution $x(t)$ of (5), say, that defined by the initial conditions $x(a) = 0$ and $x'(a) = 1$. This solution is called nonoscillatory in the interval (a, b) if it has no zeroes there; otherwise, it oscillates. There is a simple and fruitful way to study the presence/absence of zeroes of $x(\cdot)$ on (a, b), which is at the heart of the classical Sturm–Liouville theory. Namely, one introduces the polar angle $\theta(t)$ of the vector $\left[\begin{smallmatrix} x(t) \\ x'(t) \end{smallmatrix}\right]$ in the two-dimensional phase plane $\mathbb{R}^2$. It is clear that if $a < t < b$, then $x(t) = 0$ if and only if $\theta(t) = \pi/2 \bmod \pi$. Moreover, $\theta'(t) < 0$ at each zero t of $x(t)$, so we can determine the number of zeroes of $x(\cdot)$ in (a, b) by studying the evolution of $\theta(\cdot)$ there, that is, the "rotation" of $x(\cdot)$.

This simple observation does not generalize easily to the Hamiltonian system (1). It is rather straightforward to generalize the concept of zero of $x(\cdot)$: one sets $\mathbf{z} = \left[\begin{smallmatrix} \mathbf{x} \\ \mathbf{y} \end{smallmatrix}\right]$, requires that $\mathbf{x}(t) = \mathbf{0}$, and arrives at the concept of focal point, alias point of verticality. But it is not easy to extend the concept of polar angle in an appropriate way; in fact, it seems that this was only done in the 1950s and 1960s. One way is to introduce argument functions in the symplectic group, as done by Gel'fand, Lidskii, and Yakubovich. Another is to introduce the Maslov cycle and the corresponding Maslov index in the manifold of Lagrange subspaces of $\mathbb{R}^{2n}$. There is a corresponding angle, as was pointed out by Arnol'd (and by Conley in a little-known paper), which can be used to develop a Sturm–Liouville-type theory for (4). Still another method to generalize the Sturm–Liouville theory to Hamiltonian systems can be based on the polar coordinates of Barret and Reid.

A point which we will emphasize in this book is that one can study the argument functions, the index, and the polar coordinates from a dynamical point of view, more precisely, by using basic tools from topological dynamics and ergodic theory. One point of arrival in our theory is a quantity called the rotation number and its "complexification," the Floquet exponent for system (1). Using these quantities, we will connect the oscillation theory of (1) with the spectral theory of the Atkinson problem (4), much as the Sturm–Liouville theory connects the oscillation of solutions of (3) for each fixed λ to the spectral theory of (3).

Let us explain this matter in more detail. Let $\Gamma \geq 0$ be a real symmetric matrix-valued function. Consider the boundary-value problem

$$\begin{aligned} &\mathbf{z}' = \big(H(t) + \lambda\, J^{-1}\Gamma(t)\big)\,\mathbf{z}\,, \qquad \mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \in \mathbb{R}^{2n}\,, \\ &\mathbf{x}(a) = \mathbf{x}(b) = \mathbf{0}\,, \end{aligned} \tag{6}$$

where $a < b \in \mathbb{R}$. In [5] an analytic theory of the eigenvalues and eigenfunctions of (6) is worked out. Let us first try to extend that theory to the entire real axis: thus set $a = -\infty$ and $b = \infty$. One can expect that this will involve some analogue of the classical Weyl m-functions $m_{\pm}(\lambda)$ for (3), and in fact there is a rich literature concerning the "Weyl–Titchmarsh M-matrices" for (6). We will assume that $H(\cdot)$ and $\Gamma(\cdot)$ are uniformly bounded and will impose a natural "Atkinson condition" on the solutions of (5). It will then turn out that the dynamical concept of exponential dichotomy together with the above-mentioned notion of rotation number permits one to develop a satisfactory spectral theory for (6) with $a = -\infty$ and $b = \infty$. In particular, the introduction of the exponential dichotomy concept permits one to clarify the dynamical significance of the M-matrices.

To summarize what has been said so far, we will supplement the analytic methods which have been previously used to study the oscillation theory of (1) and the spectral theory of (4) with certain geometrical and dynamical techniques. The geometrical methods derive from the structure of the group of symplectic matrices and from that of the manifold of Lagrangian subspaces of $\mathbb{R}^{2n}$. Using dynamical methods, we define the rotation number and the Floquet exponent, which permit one to count the focal points of (1) and to develop the spectral theory of (4) using the exponential dichotomy concept.

The use of dynamical methods is made possible by carrying out a construction named after Bebutov, which we now explain. Begin with linear Hamiltonian differential system (1): we first view the coefficient function $H(\cdot)$ as an element of an appropriate functional space. This will often be the space of bounded continuous functions $\widetilde{H}$ from $\mathbb{R}$ to the Lie algebra of real infinitesimally symplectic matrices $\mathfrak{sp}(n,\mathbb{R}) = \{\widetilde{H} \in \mathbb{M}_{2n\times 2n}(\mathbb{R}) \mid \widetilde{H}^T J + J\widetilde{H} = 0_{2n}\}$. Next introduce the translation flow σ_t by setting $\sigma_t(\widetilde{H})(\cdot) = \widetilde{H}(\cdot + t)$ for all $t \in \mathbb{R}$. If the coefficient $H(\cdot)$ of (1) is uniformly continuous, then the closure $\mathrm{cls}\{\sigma_t(H) \mid t \in \mathbb{R}\}$ is compact (in the compact-open topology). Call the closure Ω: it is clearly invariant with respect to the translation flow. The idea now is to let H vary over Ω; to emphasize that we do not deal only with the “original” function $H(\cdot)$, we write ω to indicate a generic point of Ω. Note that each $\omega \in \Omega$ gives rise to a linear differential system of the form (1); call this system $(1)_\omega$.

At this point, one introduces the so-called cocycle obtained by considering the fundamental matrix solution of $(1)_\omega$ and letting ω run over Ω. One can now apply the Oseledets theory of the Lyapunov indices of solutions of $(1)_\omega$ ($\omega \in \Omega$). One can also apply the Sacker–Sell–Selgrade approach to the theory of exponential dichotomies. In addition, one can define the rotation number of the family of equations $(1)_\omega$. We will see that all these dynamical methods permit one to gain important insight into the oscillation theory of (1) and the spectral theory of (4).

In fact the main tool in the analysis consists in the systematic use of the rotation number, the Lyapunov index, the exponential dichotomy concept, and the Weyl matrices. These objects are also important in the discussion of two more notions which are of fundamental significance in the context of the linear Hamiltonian system (1): the property of disconjugacy, which is of basic significance in the calculus of variations, and the related property of existence of principal solutions, which in many interesting cases can be understood as a generalization to the nonuniformly hyperbolic case of the bundles provided by the existence of exponential dichotomy.

Applications to Control Theory

There are numerous applications of the oscillation theory of equation (1) to the theory of mechanical systems, to the calculus of variations, to control theory, and to other areas. We will not give an exhaustive account of these applications. But we will apply our results concerning equations (1) and (4) to certain problems in linear-quadratic (LQ) control theory. Among these are the linear-quadratic regulator problem, the Kalman–Bucy filter, the Yakubovich frequency theorem, and the question of Willems-type dissipativity in (linear) control systems. We now discuss in a bit more detail these applications to control theory.

First we recall the formulation of the LQ regulator problem. The point of departure consists of a linear control problem

$$\begin{aligned} &\mathbf{x}' = A(t)\,\mathbf{x} + B(t)\,\mathbf{u}\,, \qquad \mathbf{x} \in \mathbb{R}^n,\ \mathbf{u} \in \mathbb{R}^m\,, \\ &\mathbf{x}(0) = \mathbf{x}\,. \end{aligned} \tag{7}$$

The matrices $A(\cdot)$, $B(\cdot)$ are taken to be bounded continuous functions; the time dependence is otherwise arbitrary. Let $\tau \in (0,\infty]$ be an extended positive real number. Introduce a quadratic functional

$$\mathcal{I}_{\mathbf{x}}(\mathbf{x},\mathbf{u}) = \langle \mathbf{x}(\tau), S\mathbf{x}(\tau)\rangle + \int_0^\tau \big(\langle \mathbf{x}(t), G(t)\,\mathbf{x}(t)\rangle + \langle \mathbf{u}(t), R(t)\,\mathbf{u}(t)\rangle\big)\,dt\,.$$

where S is a symmetric positive semidefinite matrix and $G(\cdot)$, $R(\cdot)$ are bounded continuous functions such that $G^T(t) = G(t) \geq 0$ and $R^T(t) = R(t) > 0$ for all $t \in \mathbb{R}$. If the upper limit τ is finite, one speaks of a finite-horizon problem, otherwise one has an infinite-horizon problem. If $\tau = \infty$ one sets $S = 0_n$. For each fixed initial condition $\mathbf{x} \in \mathbb{R}$, one seeks a control $\mathbf{u}\colon [0, \tau] \to \mathbb{R}^m$ which, when taken together with the corresponding solution of (7), minimizes $\mathcal{I}_{\mathbf{x}}(\mathbf{x}, \mathbf{u})$.

This basic problem has been studied in detail and has been solved both when $\tau < \infty$ and when $\tau = \infty$. Our contribution is to give a solution in the infinite-horizon case $\tau = \infty$ which uses the theory of exponential dichotomies and the rotation number as applied to an appropriate linear Hamiltonian system of the form (1). In this way one obtains, among other things, detailed information concerning the regular dependence of the optimal control on parameters.

The appropriate system (1) is obtained via a formal application of the Pontryagin maximum principle. According to this principle, a minimizing control $\mathbf{u}$ must maximize the Hamiltonian

$$\mathcal{H}(t, \mathbf{x}, \mathbf{y}, \mathbf{u}) = \langle \mathbf{y}, A(t)\,\mathbf{x} + B(t)\,\mathbf{u}\rangle - \frac{1}{2}\left(\langle \mathbf{x}, G(t)\,\mathbf{x}\rangle + \langle \mathbf{u}, R(t)\,\mathbf{u}\rangle\right),$$

for each $t \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^n$, and an appropriate $\mathbf{y} \in \mathbb{R}^n$. Here $\mathbf{y}$ is interpreted as a variable dual to $\mathbf{x}$. This leads immediately to the “feedback rule”

$$\mathbf{u} = R^{-1}(t)\,B^T(t)\,\mathbf{y}\,.$$

Substituting for $\mathbf{u}$ in the Hamiltonian equations $\mathbf{x}' = \partial\mathcal{H}/\partial\mathbf{y}$, $\mathbf{y}' = -\partial\mathcal{H}/\partial\mathbf{x}$ leads to the differential system

$$\mathbf{z}' = \begin{bmatrix} A(t) & B(t)\,R^{-1}(t)\,B^T(t) \\ G(t) & -A^T(t) \end{bmatrix} \mathbf{z}\,. \tag{8}$$

Of course, (8) is a special case of (1).

We now arrive at the main point, which is that (under standard controllability and observability conditions on (7)) the system (8) admits exponential dichotomy. This is easily proved when one has available the basic facts concerning the rotation number of (8) and its relation to the existence of exponential dichotomy. Now, the existence of exponential dichotomy for (8) means that there is a linear projection $P = P^2\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ such that if $\mathbf{z} = \left[\begin{smallmatrix}\mathbf{x}\\ \mathbf{y}\end{smallmatrix}\right]$ is in the image of P, then the solution $\mathbf{z}(t)$ of (8) satisfying $\mathbf{z}(0) = \mathbf{z}$ decays exponentially as $t \to \infty$. It further turns out that $\mathbf{z}(t) = \left[\begin{smallmatrix}\mathbf{x}(t)\\ \mathbf{y}(t)\end{smallmatrix}\right] = \left[\begin{smallmatrix}\mathbf{x}(t)\\ M(t)\,\mathbf{x}(t)\end{smallmatrix}\right]$ where $\mathbf{x}(0) = \mathbf{x}$ and $M(t)$ is a function taking values in the set of negative definite symmetric $n \times n$ matrices. Set $\mathbf{u}(t) = R^{-1}(t)\,B^T(t)\,M(t)\,\mathbf{x}(t)$ and note that $\mathbf{u}(t) \to \mathbf{0}$ exponentially as $t \to \infty$. So it is not so surprising that this $\mathbf{u}$ is in fact the unique control which minimizes $\mathcal{I}_{\mathbf{x}}(\mathbf{x}, \mathbf{u})$. If one varies $\mathbf{x}$, the dichotomy projection P and the symmetric matrix-valued function $M(t)$ do not change, so in fact we have solved the LQ regulator problem.

Let us note in passing that we have also solved the feedback stabilization problem for the control system (7). In fact, set $\mathbf{u}(t) = R^{-1}(t)\,B^T(t)\,M(t)\,\mathbf{x}(t)$ as above. Note that if $\mathbf{z}(t) = \begin{bmatrix}\mathbf{x}(t)\\ \mathbf{y}(t)\end{bmatrix}$ is the solution of (8) mentioned above, then $\mathbf{x}(t)$ solves (7) with precisely this control $\mathbf{u}(t)$. Since $\mathbf{u}$ has the "feedback form" $\mathbf{u}(t) = K(t)\,\mathbf{x}(t)$ with $K(t) = R^{-1}(t)\,B^T(t)\,M(t)$, and since the linear system $\mathbf{x}' = \big(A(t) + B(t)\,R^{-1}(t)\,B^T(t)\,M(t)\big)\,\mathbf{x}$ is exponentially stable, we have "feedback stabilized" the system (7).

We can also study certain important properties of the Kalman–Bucy filter by applying our methods to an appropriate Hamiltonian system of the form (1). This is because, as Kalman and Bucy observed, the construction of their filter is closely tied to a "time-reversed" LQ regulator problem. We briefly describe the filter and the relevance of the theory of linear Hamiltonian systems in this context.

Let $\boldsymbol{\xi}(t) \in \mathbb{R}^n$ $(t \geq 0)$ denote the state of a linear system which is disturbed by a d-dimensional white noise process: thus

$$d\boldsymbol{\xi}(t) = A(t)\,\boldsymbol{\xi}(t)\,dt + S(t)\,d\mathbf{w}(t)\,. \tag{9}$$

Here $\mathbf{w}(t)$ is a d-dimensional standard Brownian motion, and equation (9) is understood to be of Itô type. The state $\boldsymbol{\xi}(t)$ can only be partially observed; it is assumed that the observation process $\boldsymbol{\eta}(t)$ satisfies the Itô equation:

$$d\boldsymbol{\eta}(t) = B(t)\,\boldsymbol{\xi}(t)\,dt + S_1(t)\,d\mathbf{w}_1(t)\,.$$

where $\mathbf{w}_1(t)$ is a second, m-dimensional Brownian motion which is independent of $\mathbf{w}(t)$. The functions A, B, S, S_1 are assumed to be continuous and bounded and to have the appropriate dimensions. It is assumed that $\boldsymbol{\eta}(0) = \mathbf{0}$ and that $\boldsymbol{\xi}(0)$ is Gaussian, which implies that $\boldsymbol{\xi}(t)$ is Gaussian for all $t \geq 0$.

Let Σ_t be the σ-algebra generated by the set $\{\boldsymbol{\eta}(r) \mid 0 \leq r \leq t\}$ of measurements up to time t. The goal is to describe an estimate $\boldsymbol{\gamma}(t)$ for $\boldsymbol{\xi}(t)$, which minimizes the mean-square error $E\{(\mathbf{x}^T(\boldsymbol{\xi}(t) - \boldsymbol{\gamma}(t)))^2\}$ for all vectors $\mathbf{x} \in \mathbb{R}^n$; here the expected value $E\{\cdot\}$ is taken over an appropriate probability space. It is well known that this best estimate is given by the conditional expectation

$$\boldsymbol{\gamma}(t) = \widehat{\boldsymbol{\xi}}(t) = E\{\boldsymbol{\xi}(t) \mid \Sigma_t\}\,.$$

To describe $\widehat{\boldsymbol{\xi}}(t)$, one introduces the error process $\tilde{\boldsymbol{\xi}}(t) = \boldsymbol{\xi}(t) - \widehat{\boldsymbol{\xi}}(t)$. It turns out that $\tilde{\boldsymbol{\xi}}(t)$ is Gaussian with mean value zero and hence is determined by its $n \times n$ covariance matrix $M(t)$. Kalman and Bucy showed that $M(t)$ satisfies a Riccati equation

$$M' = -M\,B^T(t)\,(S_1 S_1^T)^{-1}(t)\,B(t)\,M + M\,A^T(t) + A(t)\,M + (SS^T)(t)\,.$$

Now, this Riccati equation corresponds to the linear Hamiltonian system

$$\mathbf{z}' = \begin{bmatrix} -A^T(t) & B^T(t)\,(S_1 S_1^T)^{-1}(t)\,B(t) \\ (SS^T)(t) & A(t) \end{bmatrix} \mathbf{z}\,, \tag{10}$$

via the matrix change of variables $M = YX^{-1}$. It turns out that, under standard controllability conditions, the system (10) admits exponential dichotomy. This leads to the conclusion that $M(t)$ tends exponentially fast to a "nonautonomous equilibrium" $M_\infty(t)$, which essentially describes the error process $\tilde{\boldsymbol{\xi}}(t)$, and hence the signal $\boldsymbol{\xi}(t)$ if one takes the estimate $\widehat{\boldsymbol{\xi}}(t)$ to be known.

We will also apply our results concerning the oscillation theory of equation (1) and the spectral theory of the family (4) to the circle of ideas and results centered on the Yakubovich frequency theorem. This theorem was originally formulated and proved by Yakubovich for LQ control processes with periodic coefficients. We will state and prove a more general nonautonomous version of this theorem. We briefly sketch our results in this regard in the next paragraphs.

The point of departure is again the control system (7) combined with a quadratic functional

$$\widetilde{\mathcal{I}}_{\mathbf{x}}(\mathbf{x},\mathbf{u}) = \int_0^\infty \left(\langle \mathbf{x}, G(t)\,\mathbf{x}\rangle + 2\langle \mathbf{x}, g(t)\,\mathbf{u}\rangle + \langle \mathbf{u}, R(t)\,\mathbf{u}\rangle\right)\,dt\,,$$

where the functions A, B, G, g, R are assumed to be bounded and continuous and to have the appropriate dimensions. The functional $\widetilde{\mathcal{I}}_{\mathbf{x}}(\mathbf{x},\mathbf{u})$ differs from the one encountered in the context of the LQ regulator in two respects. First of all, the cross-term $\langle \mathbf{x}, g(t)\,\mathbf{u}\rangle$ is present in the integrand. Second and more importantly, though it is assumed that $G^T(t) = G(t)$ and that $R^T(t) = R(t) > 0$ for all t, it is not assumed that G is positive semidefinite for all t; indeed one is particularly interested in the case when $G(t) < 0$ ($t \in \mathbb{R}$).

We pose the problem of minimizing $\widetilde{\mathcal{I}}_{\mathbf{x}}(\mathbf{x},\mathbf{u})$ subject to (7). Since G is not assumed to be positive semidefinite, this problem need not have a solution. Nevertheless we proceed by applying the Pontryagin maximum principle in a formal way. Introduce the Hamiltonian

$$\widetilde{\mathcal{H}}(t,\mathbf{x},\mathbf{y},\mathbf{u}) = \langle \mathbf{y}, A(t)\,\mathbf{x} + B(t)\,\mathbf{u}\rangle - \frac{1}{2}\left(\langle \mathbf{x}, G(t)\,\mathbf{x}\rangle + 2\langle \mathbf{x}, g(t)\,\mathbf{u}\rangle + \langle \mathbf{u}, R(t)\,\mathbf{u}\rangle\right).$$

A minimizing control $\mathbf{u}$ (if it exists) will maximize $\widetilde{\mathcal{H}}$ for each $t \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$, and an appropriate $\mathbf{y} \in \mathbb{R}^n$. This leads to the feedback rule

$$\mathbf{u} = R^{-1}(t)\,B^T(t)\,\mathbf{y} - R^{-1}(t)\,g^T(t)\,\mathbf{x}\,,$$

and via the Hamiltonian equations $\mathbf{x}' = \partial\widetilde{\mathcal{H}}/\partial\mathbf{y}$, $\mathbf{y}' = -\partial\widetilde{\mathcal{H}}/\partial\mathbf{x}$, one is led to the differential system

$$\mathbf{z}' = H(t)\,\mathbf{z}\,, \qquad \text{with } H = \begin{bmatrix} A - B\,R^{-1}g^T & B\,R^{-1}B^T \\ G - g\,R^{-1}g^T & -A^T + g\,R^{-1}B^T \end{bmatrix}. \tag{11}$$

In the case when all the coefficients in (11) are T-periodic, Yakubovich showed that the minimization problem admits a solution if and only if (i) the system (11) has exponential dichotomy (frequency condition) and (ii) certain solutions of (11) have no focal points (nonoscillation condition). We will consider the case when A, B, G, g, R are bounded continuous functions of time and prove a satisfactory generalization of Yakubovich's theorem. It turns out that the frequency condition and the nonoscillation condition (which can be stated as above) imply that the optimal control problem can be solved for all $\mathbf{x} \in \mathbb{R}^n$. The converse statement is not quite true; as a matter of fact, and roughly speaking, the minimizing control must exhibit a uniform continuity condition in order to ensure that the frequency condition and the nonoscillation condition are valid.

The frequency theorem has many ramifications and applications, some of which will be considered in this book. Here we mention that the frequency theorem can be used to comment on the Willems concept of dissipativity in the context of control systems. This connection was pointed out and analyzed in the periodic case, by Yakubovich et al. [158]. We will discuss the connection between the frequency theorem and the Willems dissipativity concept when the relevant coefficients are aperiodic functions of time.

The main point here is to interpret the integrand of the functional $\widetilde{\mathcal{I}}_{\mathbf{x}}(\mathbf{x}, \mathbf{u})$ as a power function. To explain this, set $\mathbf{x} = \mathbf{0}$ in equation (7). Let $\mathbf{u}\colon [t_1, t_2] \to \mathbb{R}^m$ be an integrable function, and let $\mathbf{x}(t)$ be the corresponding solution of (7) with $\mathbf{x}(t_1) = \mathbf{0}$. Let us write

$$\mathcal{Q}(t, \mathbf{x}, \mathbf{u}) = \frac{1}{2} \left(\langle \mathbf{x}, G(t)\, \mathbf{x} \rangle + 2\langle \mathbf{x}, g(t)\, \mathbf{u} \rangle + \langle \mathbf{u}, R(t)\, \mathbf{u} \rangle \right) .$$

Then the net energy entering the system due to the effect of $\mathbf{u}(\cdot)$ is obtained by integrating $\mathcal{Q}(t, \mathbf{x}(t), \mathbf{u}(t))$ in the interval $[t_1, t_2]$. Now one says that the system is dissipative if

$$\int_{t_1}^{t_2} \mathcal{Q}(s, \mathbf{x}(s), \mathbf{u}(s))\, ds \geq 0$$

whenever $t_1 < t_2 \in \mathbb{R}$. That is, "energy must be expended" to move the system from its equilibrium position $\mathbf{x} = \mathbf{0}$.

The basic result which we will prove is that, modulo details, the control system determined by (7) together with $\mathcal{Q}(t, \mathbf{x}, \mathbf{u})$ is (strongly) dissipative if and only if the Hamiltonian system (11) satisfies the frequency condition and the nonoscillation condition. So the frequency theorem has deep consequences concerning the structure of LQ control processes.

Outline of the Contents

We end this introduction with a brief outline of the contents of the various chapters which will follow.

The long Chap. 1 contains a discussion of various tools from topological dynamics and from ergodic theory which will be systematically used throughout the book. We discuss the Birkhoff theorem and the Oseledets theorem, the Bebutov construction and some facts concerning flows, the Sacker–Sell–Selgrade theory of exponential dichotomies, and other matters as well.

Chapters 2 and 3 contain the basic theory of the oscillation of the solutions of (1), respectively, as well as a dynamical approach to the spectral theory of the Atkinson problem (4). In Chap. 2, we construct and discuss the rotation number for (1), which is roughly speaking "the average number of focal points" admitted by a so-called conjoined basis of solutions. This quantity can be defined in several ways, using the Gel'fand–Lidskii–Yakubovich argument functions, the Maslov index, and the Barrett–Reid polar angles. In Chap. 3 we complexify the rotation number so as to obtain the Floquet exponent, a quantity which is quite useful in the study of problem (4). We state and prove a basic result, namely, that if (4) satisfies an Atkinson condition, then the rotation number $\alpha = \alpha(\lambda)$ of (4) is constant for λ in an open subinterval $\mathcal{I} \subset \mathbb{R}$ if and only if (4) admits exponential dichotomy for all $\lambda \in \mathcal{I}$.

The Weyl M-matrices, or M-functions, arise in Chap. 3 as a tool used in the study of the spectral theory of (4) and especially in the proof of the theorem relating the constancy of the rotation number to the presence of exponential dichotomy. The M-functions are defined for nonreal values of the parameter λ. However, it is very important to understand their convergence properties in the limit as $\operatorname{Im}\lambda$ tends to zero, and Chap. 4 is dedicated to a study of this issue. In particular, we work out an extension to the Atkinson problem (4) of the classical Kotani theory, which is an important tool in the study of the refined spectral properties of the Schrödinger operator.

The notion of disconjugacy is very important in the context of the Hamiltonian linear differential system (1), because of its significance in the calculus of variations. Chapter 5 is devoted to a discussion of a generalization of the concept of disconjugacy, namely, weak disconjugacy. Under natural and mild auxiliary hypotheses, we prove the existence of a principal solution when (1) is weakly disconjugate. Our approach to the issue of (weak) disconjugacy relies on the systematic use of tools of topological dynamics; these allow a deep understanding of the conditions under which weak disconjugacy holds and also of the properties of the principal solutions.

The book concludes with Chap. 6 (the LQ regulator problem and the Kalman–Bucy filter), Chap. 7 (the nonautonomous version of the Yakubovich frequency theorem), and Chap. 8 (Willems dissipativity for LQ control processes).

Note finally that, in this book, methods and results which have been developed in the course of 100 years in the context of linear Hamiltonian systems with constant or periodic coefficients are extended to systems whose coefficients can exhibit a much more general time dependence. Indeed, techniques of topological dynamics and of ergodic theory which have been worked out in recent times permit us to apply new methods and adapt older ones to the study of a rich set of new scenarios which are not possible in the periodic case. In the end we obtain a coherent theory which has been successfully applied to a wide range of problems in the setting of nonautonomous linear Hamiltonian systems.

Firenze, Italy — Russell Johnson
Valladolid, Spain — Rafael Obaya
Valladolid, Spain — Sylvia Novo
Valladolid, Spain — Carmen Núñez
Firenze, Italy — Roberta Fabbri

Acknowledgments

The contents of this book are based on many years of research regarding diverse aspects of the theory of nonautonomous linear Hamiltonian systems. The fact that it exists is due in large part to illuminating discussions with many colleagues and to the support of relatives and friends. To all of them, many thanks.

Russell Johnson would in particular like to thank Mahesh Nerurkar for joint research activity extending over many years. A number of the results in this book can be traced to that collaboration.

Sylvia Novo would like to express her appreciation to Manuel Núñez for his role as a source of mathematical information.

Finally, Carmen Núñez wishes to thank Enrique Marty deeply, for his comprehension and continuous encouragement during the years of elaboration of the book.

Contents

Chapter 1
Nonautonomous Linear Hamiltonian Systems

This chapter is devoted to the general explanation of the framework of the analysis made in this book, and to stating the many foundational facts which will be required. With the aim of being relatively self-contained, precise references where the proofs of the stated properties can be found are included, and at the same time some proofs which the reader may consider elementary or well known, but for which it is not easy to find a completely appropriate reference in the literature, are given.

This long chapter is divided into four sections. The first presents the most fundamental notions and properties of topological dynamics and ergodic theory, including the concept and main characteristics of a skew-product flow, which are fundamental for the book.

The second section summarizes basic results concerning spaces of matrices, the Grassmannian and Lagrangian manifolds, and matrix-valued functions.

Section 1.3 is devoted to the description of the general framework of the book. Under mild conditions on the coefficient matrix, a nonautonomous linear system of ordinary differential equations defines continuous skew-product flows on the trivial and Grassmannian bundles above a compact metric space. Special attention is devoted to the Hamiltonian case, for which two special skew-product flows can be defined. For the first one, which is defined on the Lagrange bundle, the use of generalized polar coordinates simplifies the task of describing the dynamical behavior. The second one, which is closely related to the first, is defined on the bundle given by the set of symmetric matrices. It presents some interesting monotonicity properties.

The last section concerns one of the most fundamental concepts for the development of the analysis made in the book: that of exponential dichotomy, both in the general linear case and in the linear Hamiltonian case. Many of the properties ensured by its presence will be described in detail, and then applied later in the book. The closely related concept of Sacker–Sell spectrum is also discussed, and several aspects of the Sacker–Sell perturbation theory are explained. The section is

R. Johnson et al., *Nonautonomous Linear Hamiltonian Systems: Oscillation, Spectral Theory and Control*, Developments in Mathematics 36,
DOI 10.1007/978-3-319-29025-6_1

completed with the less standard analysis of the behavior of the Grassmannian flows in the presence of exponential dichotomy.

1.1 Some Fundamental Notions

The concepts and properties summarized in this section will be used often throughout the book, many times without reference to these initial pages. Suitable references for all these notions include Nemytskii and Stepanov [110], Ellis [41], Sacker and Sell [133], Cornfeld et al. [35], Walters [148], Mañé [99], and Rudin [128, 129].

1.1.1 Basic Concepts and Properties of Topological Dynamics

Let Ω be a locally compact Hausdorff topological space. Let Σ_Ω and $\Sigma_\mathbb{R}$ represent the Borel sigma-algebras of Ω and $\mathbb{R}$, and let $\Sigma_* = \Sigma_\mathbb{R} \times \Sigma_\Omega$ be the product sigma-algebra; i.e. the intersection of all the sigma-algebras on $\mathbb{R} \times \Omega$ containing the sets $\mathcal{I} \times \mathcal{A}$ for $\mathcal{I} \in \Sigma_\mathbb{R}$ and $\mathcal{A} \in \Sigma_\Omega$. Mild conditions on Ω ensure that Σ_* agrees with the Borel sigma-algebra of $\mathbb{R} \times \Omega$: it is enough to assume that Ω admits a countable basis of open sets (see e.g. Proposition 7.6.2 of Cohn [30]).

It will be convenient to work under the hypothesis that Σ_* is indeed the Borel sigma-algebra of $\mathbb{R}\times\Omega$. So, throughout Sect. 1.1, Ω will represent a locally compact Hausdorff topological space which admits a countable basis of open sets. In fact, throughout the book, any flow will be defined on a set which satisfies, at a minimum, these conditions. Some of the results explained in this section require Ω to be a compact metric space, but this hypothesis will be specified whenever it is assumed.

A map $\sigma\colon \mathbb{R} \times \Omega \to \Omega$ is *Borel measurable* if $\sigma^{-1}(\mathcal{A}) \in \Sigma_*$ for all $\mathcal{A} \in \Sigma_\Omega$. A *global real Borel measurable flow* on Ω is a Borel measurable map $\sigma\colon \mathbb{R} \times \Omega \to \Omega$ such that $\sigma_0 = \mathrm{Id}_\Omega$ and $\sigma_{t+s} = \sigma_t \circ \sigma_s$ for all $s, t \in \mathbb{R}$, where $\sigma_t\colon \Omega \to \Omega,\ \omega \mapsto \sigma(t,\omega)$. The flow is *continuous* if σ satisfies the stronger condition of being a continuous map, in which case each map σ_t is a homeomorphism on Ω with inverse σ_{-t}. The notation (Ω, σ) will be frequently used to represent a real global flow on Ω, and the words *real* and *global* will be omitted when no confusion arises.

The *orbit* of a point $\omega \in \Omega$ is the set $\{\sigma_t(\omega) \mid t \in \mathbb{R}\}$, and its *positive* (resp. *negative*) *semiorbit* is $\{\sigma_t(\omega) \mid t \in \mathbb{R}_+\}$, where $\mathbb{R}_+ = \{t \in \mathbb{R} \mid t \ge 0\}$ (resp. $\{\sigma_t(\omega) \mid t \in \mathbb{R}_-\}$, where $\mathbb{R}_- = \{t \in \mathbb{R} \mid t \le 0\}$).

Given a Borel measurable flow (Ω, σ), a Borel subset $\mathcal{A} \subseteq \Omega$ (i.e. an element $\mathcal{A}$ of Σ_Ω) is σ*-invariant* (resp. *positively* or *negatively* σ*-invariant*) if $\sigma_t(\mathcal{A}) = \mathcal{A}$ for all $t \in \mathbb{R}$ (resp. $t \in \mathbb{R}_+$ or $t \in \mathbb{R}_-$). Let $\mathbb{Y}$ be a topological space. If Σ is a sigma-algebra on Ω containing the Borel sets, a map $f\colon \Omega \to \mathbb{Y}$ is Σ*-measurable* if $f^{-1}(\mathcal{B}) \in \Sigma$ for every Borel subset $\mathcal{B} \subseteq \mathbb{Y}$; and f is *Borel measurable* when it is Σ_Ω-measurable. A Borel measurable function $f\colon \Omega \to \mathbb{Y}$ is σ*-invariant* if $f(\sigma_t(\omega)) =$

$f(\omega)$ for all $\omega \in \Omega$ and $t \in \mathbb{R}$. It is obvious that a Borel subset $\mathcal{A}$ is σ-invariant if and only if its characteristic function $\chi_{\mathcal{A}}$ is σ-invariant.

If Σ is a sigma-algebra containing the Borel sets, the concepts of σ-invariant set $\mathcal{A} \in \Sigma$ and σ-invariant Σ-measurable map $f : \Omega \to \mathbb{Y}$ are defined analogously. Note that in fact this concept of invariance can be extended to any set or function, since it does not depend on measurability.

All these definitions of *σ-invariance* correspond to *strict σ-invariance*, although the word *strict* will be almost always omitted. A less restrictive definition of invariance, depending on a fixed measure, is given in Sect. 1.1.2.

The flow is *local* if the map σ is defined, Borel measurable, and satisfies the two initially required properties on an open subset $\mathcal{O} \subseteq \mathbb{R} \times \Omega$ containing $\{0\} \times \Omega$. Define $\mathcal{O}_\omega = \{t \in \mathbb{R} \mid (t, \omega) \in \mathcal{O}\}$ for $\omega \in \Omega$. The *orbit* of the point ω for a local flow (Ω, σ) is $\{\sigma_t(\omega) \mid t \in \mathcal{O}_\omega\}$, and it is *globally defined* if $\mathcal{O}_\omega = \mathbb{R}$. The *positive* (resp. *negative*) *semiorbit* of a point ω is the set $\{\sigma_t(\omega) \mid t \in \mathcal{O}_\omega \cap \mathbb{R}_+\}$ (resp. $\{\sigma_t(\omega) \mid t \in \mathcal{O}_\omega \cap \mathbb{R}_-\}$, and it is *globally defined* if $\mathcal{O}_\omega \cap \mathbb{R}_+ = \mathbb{R}_+$ (resp. $\mathcal{O}_\omega \cap \mathbb{R}_- = \mathbb{R}_-$). A (in general Borel) subset $\mathcal{A} \subseteq \Omega$ is *σ-invariant* (resp. *positively* or *negatively σ-invariant*) if it is composed of globally defined orbits (resp. globally defined positive or negative semiorbits).

Finally, replacing $\mathbb{R}$ by $\mathbb{R}_+$ (resp. by $\mathbb{R}_-$) provides the definition of a (global or local) real positive (resp. negative) *semiflow* on Ω. The definitions of positive (resp. negative) semiorbit and (strict) invariance are the obvious ones.

For the remaining definitions and properties discussed in this section, the flow σ is assumed to be continuous.

A compact σ-invariant subset $\mathcal{M} \subseteq \Omega$ is *minimal* if it does not contain properly any other such set; or, equivalently, if each of its positive or negative semiorbits is dense in it. The flow (Ω, σ) is *minimal* or *recurrent* if Ω itself is minimal, which obviously requires Ω to be compact. Note that Zorn's lemma ensures that, if Ω is compact, then it contains at least one minimal subset.

Suppose that the positive semiorbit of a point ω_0 for such a flow is relatively compact. Then the *omega-limit set* of the point (or of its positive semiorbit) is given by those points $\omega \in \Omega$ such that $\omega = \lim_{k\to\infty} \sigma(t_k, \omega_0)$ for some sequence $(t_k) \uparrow \infty$. The omega-limit set is nonempty, compact, connected, and σ-invariant. The concept of *alpha-limit set* is analogous, working now with a negative semiorbit and with sequences $(t_k) \downarrow -\infty$. Clearly, a minimal subset of Ω is the omega-limit set and the alpha-limit set of each of its elements.

Finally, assume in addition that Ω is a compact metric space, and let d_Ω represent the distance on Ω. The flow (Ω, σ) is *chain recurrent* if given $\varepsilon > 0$, $t_0 > 0$, and points $\omega, \tilde{\omega} \in \Omega$, there exist points $\omega = \omega_0, \omega_1, \ldots, \omega_m = \tilde{\omega}$ of Ω and real numbers $t_1 > t_0, \ldots, t_m > t_0$ such that $d_\Omega(\sigma_{t_i}(\omega_i), \omega_{i+1}) < \varepsilon$ for $i = 0, \ldots, m-1$. It is easy to check that minimality implies chain recurrence: just take $\omega_0 = \omega$ and $\omega_1 = \tilde{\omega}$ and keep in mind that the positive semiorbit of ω is dense in Ω. It is also easy to check that if (Ω, σ) is chain recurrent, then the set Ω is connected.

1.1.2 Basic Concepts and Properties of Measure Theory

Unless otherwise indicated, any measure appearing in the book is a positive normalized regular Borel measure. Given such a measure m, let Σ_m be the m-completion of the Borel sigma-algebra (see e.g. Theorem 1.36 of [128]), and represent with the same symbol m the extension of the initial measure to Σ_m. As usual, the notation "m-a.e." means *almost everywhere with respect to m*; "for m-a.e. $\omega \in \Omega$" means *for almost every* $\omega \in \Omega$; and $L^1(\Omega, m)$ represents the quotient set of Σ_m-measurable functions $f: \Omega \to \mathbb{R}$ with $\int_\Omega |f(\omega)|\, dm < \infty$ (so that two real functions represent the same class if they are m-a.e. equal, in which case they are the same element of $L^1(\Omega, m)$). See Sect. 1.2.4 for the general definitions of L^p spaces of matrix-valued functions on Ω.

Let m be a measure on Ω. Then m is *σ-invariant* if $m(\sigma_t(\mathcal{A})) = m(\mathcal{A})$ for every Borel subset $\mathcal{A} \subseteq \Omega$ and all $t \in \mathbb{R}$, which ensures the same property for every $\mathcal{A} \in \Sigma_m$. A Σ_m-measurable map $f: \Omega \to \mathbb{Y}$ (for a topological space $\mathbb{Y}$) is *σ-invariant with respect to m* if, for all $t \in \mathbb{R}$, $f(\sigma_t(\omega)) = f(\omega)$ m-a.e. And a subset $\mathcal{A} \in \Sigma_m$ is *σ-invariant with respect to m* if $\chi_{\mathcal{A}}$ has this property.

The expression "σ-invariant" (for sets, measures, or functions) will often be changed to "invariant" throughout the book, since in most cases no confusion arises.

Proposition 1.2 shows the relation between these concepts of σ-invariance with respect to m and the (strict) ones given in the previous section: it proves that, when moving for instance in the quotient space $L^1(\Omega, m)$, one can always consider that a "σ-invariant function" satisfies the "strict" definition. More information in this regard will be added in Proposition 1.5.

Remark 1.1 Recall that any Σ_m-measurable function $f: \Omega \to \mathbb{K}$, for $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, agrees m-a.e. with a Borel measurable one (see [128], Lemma 1 of Theorem 8.12). In addition, if Σ is any sigma-algebra containing the Borel sets, and if a sequence $(f_n: \Omega \to \mathbb{K})$ of Σ-measurable functions converges everywhere to a function f, then f is Σ-measurable (see [128], Theorem 1.14). And, as a consequence of this last result, if $(f_n: \Omega \to \mathbb{K})$ is a sequence of Σ_m-measurable functions which converges m-a.e. to a function f, then f is Σ_m-measurable.

Proposition 1.2 *Let (Ω, σ) be a Borel measurable flow, and let m be a σ-invariant measure on Ω.*

(i) *Let the Σ_m-measurable function $f: \Omega \to \mathbb{K}$ be σ-invariant with respect to m. Then there exists a Σ_m-measurable function $f^*: \Omega \to \mathbb{K}$ which is (strictly) σ-invariant such that $f = f^*$ m-a.e.*
(ii) *Let the set $\mathcal{A} \in \Sigma_m$ be σ-invariant with respect to m. Then there exists a (strictly) σ-invariant set $\mathcal{A}^* \in \Sigma_m$ such that $\chi_{\mathcal{A}} = \chi_{\mathcal{A}^*}$ m-a.e.*

Proof

(i) The proof of this property is carried out in Lemma 1 of Chapter 1.2 of [35], and included here for the reader's convenience. It follows from Remark 1.1 that there is no loss of generality in assuming that f is Borel measurable. Define the sets $\mathcal{N} = \{(t,\omega) \in \mathbb{R} \times \Omega \mid f(\omega) \neq f(\sigma_t(\omega))\}$, and note that the hypotheses on σ ensure that this set belongs to $\Sigma_* = \Sigma_{\mathbb{R}} \times \Sigma_\Omega$, since the maps $\mathbb{R} \times \Omega \to \mathbb{R}$, $(t,\omega) \mapsto f(\omega)$ and $\mathbb{R} \times \Omega \to \mathbb{R}$, $(t,\omega) \mapsto f(\sigma_t(\omega))$ are Σ_*-measurable. Define now $\mathcal{N}_t = \{\omega \in \Omega \mid (t,\omega) \in \mathcal{N}\}$ for $t \in \mathbb{R}$, and $\mathcal{N}_\omega = \{t \in \mathbb{R} \mid (t,\omega) \in \mathcal{N}\}$ for $\omega \in \Omega$, and note that $\mathcal{N}_t \in \Sigma_\Omega$ for all $t \in \mathbb{R}$ and $\mathcal{N}_\omega \in \Sigma_{\mathbb{R}}$ for all $\omega \in \Omega$ (see Theorem 8.2 of [128]). By definition of σ-invariance with respect to m, $m(\mathcal{N}_t) = 0$ for all $t \in \mathbb{R}$. Define μ as the product measure of m and l on $\Omega \times \mathbb{R}$, where l is the Lebesgue measure on $\mathbb{R}$. Fubini's theorem (see Theorem 8.8 of [128]) ensures that the maps $\omega \mapsto l(\mathcal{N}_\omega)$ and $t \mapsto m(\mathcal{N}_t)$ are Borel, and that $\mu(\mathcal{N}) = \int_\Omega l(\mathcal{N}_\omega)\,dm = \int_{\mathbb{R}} m(\mathcal{N}_t)\,dl = 0$. Therefore the subset $\Omega_f \subseteq \Omega$ of points ω with $l(\mathcal{N}_\omega) = 0$ is Borel, and $m(\Omega_f) = 1$. Suppose that ω and $\sigma_t(\omega)$ belong to Ω_f for a pair $(t,\omega) \in \mathbb{R} \times \Omega$. Then $f(\omega) = f(\sigma_t(\omega))$. In order to prove this assertion, take $s \in \mathbb{R} - \mathcal{N}_{\sigma_t(\omega)}$ such that $s + t \in \mathbb{R} - \mathcal{N}_\omega$, and note that $f(\sigma_t(\omega)) = f(\sigma_s(\sigma_t(\omega))) = f(\sigma_{s+t}(\omega)) = f(\omega)$. Now define

$$f^*(\omega) = \begin{cases} f(\omega) & \text{if there exists } t \in \mathbb{R} \text{ with } \sigma_t(\omega) \in \Omega_f\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

which is Σ_m-measurable, since it agrees with f at least on Ω_f (and hence m-a.e.), and which is σ-invariant in the classical sense.

(ii) Let $g = \chi^*_{\mathcal{A}}$ be the σ-invariant function associated to $\chi_{\mathcal{A}}$ by (i). Then the set $\mathcal{B} = \{\omega \in \Omega \mid g(\omega) \in \{0,1\}\} = 1$ belongs to Σ_m, is σ-invariant, and has full measure for m: $m(\mathcal{B}) = 1$. The set $\mathcal{A}^* = \{\omega \in \Omega \mid g(\omega) = 1\} \subseteq \mathcal{B}$ also belongs to Σ_m and is σ-invariant. In addition, $g(\omega) = \chi_{\mathcal{A}^*}(\omega)$ for all $\omega \in \mathcal{B}$, so that $\chi_{\mathcal{A}} = \chi_{\mathcal{A}^*}$ m-a.e., as asserted.

One of the most fundamental results in measure theory is the Birkhoff ergodic theorem, one of whose simplest versions is now recalled.

Theorem 1.3 *Let (Ω,σ) and m be a Borel measurable flow and a σ-invariant measure on Ω. Given $f \in L^1(\Omega,m)$, there exists a (strictly) σ-invariant set $\Omega_f \in \Sigma_m$ with $m(\Omega_f) = 1$ such that, for all $\omega \in \Omega_f$, the limits*

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f(\sigma_s(\omega))\,ds = \lim_{t\to\infty} \frac{1}{2t}\int_{-t}^t f(\sigma_s(\omega))\,ds = \lim_{t\to-\infty} \frac{-1}{t}\int_t^0 f(\sigma_s(\omega))\,ds$$

exist, agree, and take on a real value $\tilde f(\omega)$. In addition, $\tilde f(\sigma_t(\omega)) = \tilde f(\omega)$ for all $\omega \in \Omega_f$ and $t \in \mathbb{R}$, $\tilde f$ belongs to $L^1(\Omega,m)$, and $\int_\Omega \tilde f(\omega)\,dm = \int_\Omega f(\omega)\,dm$.

Its proof in the case of a discrete flow (given by the iteration of an automorphism on Ω) can be found, for example, in Section II.1 of [99]. The procedure to deduce the result for a real flow from the discrete case is standard: define the automorphism $T(\omega) = \sigma(1,\omega)$ and, given $f \in L^1(\Omega,m)$, define $F(\omega) = \int_0^1 f(\sigma_s(\omega))\,ds$; then, Fubini's theorem ensures that $F \in L^1(\Omega,m)$, and the application of the discrete version of the theorem to this setting provides the sets Ω_f and the function $\tilde f$ satisfying the theses of the real version. The details are left to the reader.

Note that the function $\tilde{f}$ provided by the previous theorem can be considered to be σ-invariant in the strict sense: just define it to be 0 outside Ω_f. Note also that the set Ω_f contains a Borel subset with measure 1, which is clearly σ-invariant with respect to m. But in fact this Borel subset of Ω_f can be taken as a (strictly) σ-invariant set, as Proposition 1.5(i) below proves. Therefore, there is no loss of generality in assuming that the set Ω_f itself is Borel.

The following result, whose proof is included for completeness, will be required in Chap. 4. The notation $g\colon \Omega \to [0,\infty]$ is used for *extended-real* functions (which can take the value ∞), and the concept of Σ_m-measurability for such a function is clear.

Proposition 1.4 *Let (Ω,σ) be a Borel measurable flow, and let m be a σ-invariant measure on Ω. Let $f\colon \Omega \to [0,\infty)$ be a Σ_m-measurable function. Then, there exists a (strictly) σ-invariant set $\Omega_f \in \Sigma_m$ with $m(\Omega_f) = 1$ such that, for all $\omega \in \Omega_f$, the limits*

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f(\sigma_s(\omega))\,ds = \lim_{t\to\infty}\frac{1}{2t}\int_{-t}^t f(\sigma_s(\omega))\,ds = \lim_{t\to-\infty}\frac{-1}{t}\int_t^0 f(\sigma_s(\omega))\,ds$$

exist, agree, and take a value $\tilde{f}(\omega) \in \mathbb{R}\cup\{\infty\}$. In addition, the extended-real function $\tilde{f}\colon \Omega \to [0,\infty]$ is Σ_m-measurable, and it satisfies $\tilde{f}(\sigma_t(\omega)) = \tilde{f}(\omega)$ for all $\omega \in \Omega_f$ and $t \in \mathbb{R}$, and $\int_\Omega \tilde{f}(\omega)\,dm = \int_\Omega f(\omega)\,dm$.

Proof Let $h\colon \Omega \to [0,\infty)$ be a Σ_m-measurable function. For each $k \in \mathbb{N}$, define $h_k = \min(h,k)$, which obviously belongs to $L^1(\Omega,m)$. Hence there exists a function $\tilde{h}_k \in L^1(\Omega,m)$ and a set $\Omega_{h_k} \in \Sigma_m$ with $m(\Omega_{h_k}) = 1$ satisfying the theses of Theorem 1.3. Define $\Omega_h^* = \cap_{k\in\mathbb{N}}\Omega_{h_k}$, which belongs to Σ_m, is σ-invariant, and has full measure for m. Note that the nondecreasing sequence $(h_k(\omega))$ converges to $h(\omega)$ for all $\omega \in \Omega_h^*$, and define $h^*(\omega) \in [0,\infty]$ as the limit of the nondecreasing sequence of σ-invariant functions $(\tilde{h}_k(\omega))$, also for $\omega \in \Omega_h^*$. Then, h^* is Σ_m-measurable (see Remark 1.1) and σ-invariant. In addition, if $h^* \in L^1(\Omega,m)$, then $h \in L^1(\Omega,m)$: apply the Lebesgue monotone convergence theorem and the Birkhoff Theorem 1.3 to get $0 \le \int_\Omega h(\omega)\,dm = \lim_{k\to\infty}\int_\Omega h_k(\omega)\,dm = \lim_{k\to\infty}\int_\Omega \tilde{h}_k(\omega)\,dm = \int_\Omega h^*(\omega)\,dm < \infty$.

Returning to the function f of the statement, note that if $f \in L^1(\Omega,m)$, the assertions follow from Theorem 1.3. Assume hence that $\int_\Omega f(\omega)\,dm = \infty$, and associate to it the sequences (f_k) and $(\tilde{f}_k)$, the set Ω_f^*, and the function f^*, as above. Therefore, $f^* \notin L^1(\Omega,m)$. Clearly, the sets

$$\mathcal{A} = \{\omega \in \Omega_f^* \mid f^*(\omega) = \infty\}\,,$$

$$\mathcal{A}_j = \{\omega \in \Omega_f^* \mid j \le f^*(\omega) < j+1\} \quad \text{for } j \ge 0$$

belong to Σ_m, are σ-invariant and disjoint, and satisfy $\Omega_f^* = \mathcal{A} \cup (\cup_{j=0}^{\infty} \mathcal{A}_j)$. Then, if $\omega \in \mathcal{A}$,

$$\begin{aligned}\liminf_{t\to\infty} \frac{1}{t}\int_0^t f(\sigma_s(\omega))\, ds &\geq \sup_{k\in\mathbb{N}} \lim_{t\to\infty} \frac{1}{t}\int_0^t f_k(\sigma_s(\omega))\, ds \\ &= \sup_{k\in\mathbb{N}} \tilde{f}_k(\omega)\, ds = f^*(\omega) = \infty\,,\end{aligned}$$

so that there exists $\lim_{t\to\infty}(1/t)\int_0^t f(\sigma_s(\omega))\, ds = f^*(\omega) = \infty$. The same property holds for the other two limits of the proposition. Now define

$$g = \sum_{j=0}^{\infty} \frac{1}{j+1}\, \chi_{\mathcal{A}_j} f$$

on Ω_f^*, note that it is Σ_m-measurable, and associate to it the sequences (g_k), $(\tilde{g}_k)$, and the set $\Omega_g^* \subseteq \Omega_f^*$, as at the beginning of the proof. Fix any $k \in \mathbb{N}$ and any $\omega \in \Omega_g^*$ outside $\mathcal{A}$, and take the unique $j \in \mathbb{N}$ such that $\omega \in \mathcal{A}_j \cap \Omega_g^*$. Then $g(\omega) = (1/j+1) f(\omega)$, and hence

$$g_k(\omega) = \frac{1}{j+1}\,\min(f(\omega), k(j+1)) = \frac{1}{j+1} f_{k(j+1)}(\omega)\,.$$

Since $\sigma_s(\omega) \in \mathcal{A}_j \cap \Omega_g^*$ for all $s \in \mathbb{R}$,

$$\begin{aligned}\tilde{g}_k(\omega) &= \lim_{t\to\infty} \frac{1}{t}\int_0^t g_k(\sigma_s(\omega))\, ds = \frac{1}{j+1}\lim_{t\to\infty}\frac{1}{t}\int_0^t f_{k(j+1)}(\sigma_s(\omega))\, ds \\ &= \frac{1}{j+1}\tilde{f}_{k(j+1)}(\omega) \leq \frac{1}{j+1} f^*(\omega) \leq 1\end{aligned}$$

for all $k \in \mathbb{N}$. Note that g_k vanishes outside $\cup_{j=1}^{\infty}\mathcal{A}_j$. Hence $\int_\Omega g_k(\omega)\, dm = \int_\Omega \tilde{g}_k(\omega)\, dm \leq 1$, so that the Lebesgue dominated convergence theorem ensures that $g \in L^1(\Omega, m_0)$. Let $\tilde{g}$ and $\Omega_g \subseteq \Omega_f^*$ be the σ-invariant function and subset associated to g by Theorem 1.3, with $m(\Omega_g) = 1$. Then for all ω in the σ-invariant set $\mathcal{A}_j \cap \Omega_g, f(\omega) = (j+1)\, g(\omega)$ and hence

$$\begin{aligned}\lim_{t\to\infty} \frac{1}{t}\int_0^t f(\sigma_s(\omega))\, ds &= \lim_{t\to\infty}\frac{1}{2t}\int_{-t}^t f(\sigma_s(\omega))\, ds \\ &= \lim_{t\to-\infty}\frac{-1}{t}\int_t^0 f(\sigma_s(\omega))\, ds = (j+1)\,\tilde{g}(\omega) = (j+1)\,\chi_{\mathcal{A}_j}\tilde{g}(\omega)\,.\end{aligned}$$

Define $\Omega_f = \mathcal{A} \cup \Big((\cup_{j=0}^{\infty} \mathcal{A}_j) \cap \Omega_g \Big)$, and note that it belongs to Σ_m and satisfies $m(\Omega_f) = 1$. This σ-invariant set and the Σ_m-measurable and σ-invariant function

$$\tilde{f} = \begin{cases} f^*(\omega) & \text{if } \omega \in \mathcal{A} \\ \displaystyle\sum_{j=0}^{\infty} (j+1)\chi_{\mathcal{A}_j}\tilde{g} & \text{if } \omega \in (\cup_{j=0}^{\infty}\mathcal{A}_j) \cap \Omega_g \end{cases} \tag{1.1}$$

satisfy the statements regarding the limits. In addition, for all $\omega \in \Omega_f$,

$$\tilde{f}(\omega) = \lim_{t\to\infty} \frac{1}{t}\int_0^t f(\sigma_s(\omega))\,ds \geq \lim_{t\to\infty} \frac{1}{t}\int_0^t f_n(\sigma_s(\omega))\,ds = \tilde{f}_n(\omega)\,,$$

so that $\tilde{f}(\omega) \geq f^*(\omega)$ on Ω_f. Hence, $\int_\Omega \tilde{f}(\omega)\,dm \geq \int_\Omega f^*(\omega)\,dm = \infty = \int_\Omega f(\omega)\,dm$, which completes the proof.

As in the case of Theorem 1.3, the function $\tilde{f}$ provided by Proposition 1.4 can be considered to be σ-invariant in the strict sense, and Proposition 1.5(i), which is proved immediately below, ensures that the set Ω_f contains a Borel subset with measure 1 which is σ-invariant with respect to m.

Proposition 1.5 *Let (Ω, σ) be a Borel measurable flow, and let m be a σ-invariant measure on Ω.*

(i) *Let $\mathcal{A} \in \Sigma_m$ be a (strictly) σ-invariant set with $m(\mathcal{A}) = 1$. Then $\mathcal{A}$ contains a (strictly) σ-invariant Borel set $\mathcal{B}$ with $m(\mathcal{B}) = 1$.*
(ii) *Let $f\colon \Omega \to \mathbb{R}$ be Σ_m-measurable and σ-invariant with respect to m_0. Then there exists $g\colon \Omega \to \mathbb{R}$ which is Borel and (strictly) σ-invariant such that $g = f$ m-a.e.*

Proof

(i) It suffices to prove that for all $n \in \mathbb{N}$ there exists a σ-invariant Borel set $\mathcal{B}_n \subseteq \mathcal{A}$ with $m(\mathcal{B}_n) \geq m(\mathcal{A}) - 1/n$, and then take $\mathcal{B} = \cup_{n\geq 1}\mathcal{B}_n$.

Fix $n \in \mathbb{N}$, and note that the regularity of the measure m implies the existence of a compact set $\mathcal{K}_n \subseteq \mathcal{A}$ with $m(\mathcal{A} - \mathcal{K}_n) \leq 1/n$. The Borel measurability of the flow ensures that the map $\mathbb{R} \times \Omega \to \mathbb{R}\,,\ (t,\omega) \mapsto \chi_{\mathcal{K}_n}(\sigma(t,\omega))$ is Borel measurable, and hence Fubini's theorem guarantees that the maps $h_n^j\colon \Omega \to \mathbb{R}$ given by

$$h_n^j(\omega) = \sum_{i=-j}^{j} \frac{1}{|i|^2+1}\int_j^{j+1} \chi_{\mathcal{K}_n}(\sigma_t(\omega))\,dt$$

are Borel measurable (see e.g. Theorem 8.8 of [128]). Clearly, $h_n^j \leq h_n^{j+1}$, so that the limit $h_n(\omega) = \lim_{j\to\infty} h_n^j(\omega)$ exists for all $\omega \in \Omega$, and the (bounded) function h_n is Borel measurable. Define

$$\mathcal{B}_n = \{\omega \in \Omega \mid \sigma_t(\omega) \in \mathcal{K}_n \text{ for all } t \text{ in a set of positive Lebesgue measure}\}\,,$$

which is contained in $\mathcal{A}$ and is Borel, since it agrees with $h_n^{-1}((0,\infty))$. Clearly, $\mathcal{B}_n$ is (strictly) σ-invariant. The Birkhoff Theorem 1.3 ensures that the limits $l_n(\omega) = \lim_{t\to\infty}(1/2t)\int_{-t}^{t}\chi_{\mathcal{K}_n}(\omega{\cdot}s)\,ds$ exist for all ω in a σ-invariant subset $\Omega_n \in \Sigma_m$ with $m(\Omega_n) = 1$, and that the function l_n is Σ_m-measurable and σ-invariant in Ω_n. Now write $\Omega_n = \Omega_n^0 \cup \Omega_n^+$, where $\Omega_n^0 = \{\omega \in \Omega_n \mid l_n(\omega) = 0\}$ and $\Omega_n^+ = \{\omega \in \Omega_n \mid l_n(\omega) > 0\}$, and note that these sets belong to Σ_m and are σ-invariant. Applying again the Birkhoff Theorem 1.3 to the function $\chi_{\mathcal{K}_n\cap\Omega_n^0} \le \chi_{\mathcal{K}_n}$ one proves that $m(\mathcal{K}_n \cap \Omega_n^0) = \int_\Omega \chi_{\mathcal{K}_n\cap\Omega_n^0}(\omega)\,dm = 0$. On the other hand, it is clear that $\Omega_n^+ \subseteq \mathcal{B}_n$. Since

$$m(\Omega_n^0) = m(\mathcal{K}_n \cap \Omega_n^0) + m((\Omega - \mathcal{K}_n)\cap\Omega_n^0) \le m(\Omega - \mathcal{K}_n) \le \frac{1}{n}$$

and

$$m(\mathcal{B}_n) \ge m(\Omega_n^+) = 1 - m(\Omega_n^0) \ge 1 - \frac{1}{n}\,,$$

the set $\mathcal{B}_n$ satisfies the required conditions.

(ii) Remark 1.1 and the definition of σ-invariance with respect to m show that there is no loss of generality in assuming that the function f is Borel measurable. Note also that $f = f^+ - f^-$ for $f^+ = \max(f,0)$ and $f^- = -\min(f,0)$, which are Borel measurable, σ-invariant with respect to m, and nonnegative; hence it is enough to prove the result for $f \ge 0$. Now, on the one hand, repeating the argument of Proposition 1.2 one can check that the Borel set $\mathcal{N}_\omega = \{t \in \mathbb{R} \mid f(\sigma_t(\omega)) \ne f(\omega)\}$ has zero Lebesgue measure for all the points ω in a Borel set $\Omega_0 \subseteq \Omega$ with $m(\Omega_0) = 1$. And, on the other hand, Proposition 1.4 provides a σ-invariant set $\Omega_f \in \Sigma_m$ with $m(\Omega_f) = 1$ and an extended-real Σ_m-measurable σ-invariant function $\tilde f$ such that $\tilde f(\omega) = \lim_{t\to\infty}(1/t)\int_0^t f(\sigma_s(\omega))\,ds$ exists for any $\omega \in \Omega_f$. Note that if $\omega \in \Omega_0 \cap \Omega_f$, then $\tilde f(\omega)$ exists and agrees with $f(\omega)$, so that the (σ-invariant) function $\tilde f$ takes real values in a σ-invariant and Σ_m-measurable set $\widetilde{\Omega}_f \subset \Omega_f$ with $m(\widetilde{\Omega}_f) = 1$. The already verified point (i) guarantees the existence of a Borel σ-invariant set $\mathcal{B} \subseteq \widetilde{\Omega}_f$ with $m(\mathcal{B}) = 1$. Define $g = \tilde f\chi_{\mathcal{B}}$, and note that $g(\omega) = \lim_{t\to\infty}(1/t)\int_0^t f(\sigma_s(\omega))\,\chi_{\mathcal{B}}(\sigma_s(\omega))\,ds$ for any $\omega \in \Omega$. A new application of Fubini's theorem ensures that the map $\omega \mapsto (1/t)\int_0^t f(\sigma_s(\omega))\,\chi_{\mathcal{B}}(\sigma_s(\omega))\,ds$ is Borel for any $t \in \mathbb{R}$, so that g is Borel (see Remark 1.1). Clearly, it is also σ-invariant. And it agrees with f on $\mathcal{B} \cap \Omega_0$, which completes the proof.

A (positive normalized regular Borel) measure m is *σ-ergodic* if it is invariant and, in addition, any σ-invariant set has measure 0 or 1. The following fundamental property will often be applied in combination with Theorem 1.3, which associates a σ-invariant function $\tilde{f} \in L^1(\Omega, m)$ to each $f \in L^1(\Omega, m)$.

Theorem 1.6 *Let (Ω, σ) and m be a Borel measurable flow and a σ-invariant measure on Ω. The measure m is σ-ergodic if and only if every σ-invariant function $f \in L^1(\Omega, m)$ is constant m-a.e. In other words, if and only if for every $f \in L^1(\Omega, m)$ there exists a (strictly) σ-invariant set $\Omega_f \in \Sigma_m$ with $m(\Omega_f) = 1$ such that $f(\omega_0) = \int_\Omega f(\omega)\, d\mu$ for every $\omega_0 \in \Omega_f$.*

The direct implication can be proved as (1)⇒(2) in Proposition II.2.1 of [99]. As for the converse implication: if $m(\mathcal{A}) \in (0, 1)$ for a σ-invariant subset $\mathcal{A} \subseteq \Omega$, then $\chi_{\mathcal{A}}$ is a nonconstant σ-invariant integrable function. Note once more that the sets Ω_f of the previous statement can be assumed to be Borel.

The following basic characterization of invariance will be useful in the proofs of several results.

Proposition 1.7 *Let (Ω, σ) be a Borel measurable flow, and let m be a measure on Ω. The following statements are equivalent:*

(1) *m is σ-invariant;*
(2) $\int_\Omega f(\omega)\, dm = \int_\Omega f(\sigma_t(\omega))\, dm$ *for all $f \in L^1(\Omega, m)$ and all $t \in \mathbb{R}$;*
(3) $\int_\Omega f(\omega)\, dm = \int_\Omega f(\sigma_t(\omega))\, dm$ *for all $f \in C(\Omega, \mathbb{R})$ and all $t \in \mathbb{R}$.*

Proof (1)⇒(2) If the measure is invariant, then $\int_\Omega s(\omega)\, dm = \int_\Omega s(\sigma_t(\omega))\, dm$ for every simple function s. Take a nonnegative function $f \in L^1(\Omega, m)$ and choose a nondecreasing sequence (s_k) of nonnegative simple functions such that $f(\omega) = \lim_{m\to\infty} s_k(\omega)$ for all $\omega \in \Omega$ (see [128], Theorem 1.17). Hence $f(\sigma_t(\omega)) = \lim_{m\to\infty} s_k(\sigma_t(\omega))$ for all $\omega \in \Omega$ and $t \in \mathbb{R}$. Now apply the Lebesgue monotone convergence theorem in order to prove that

$$\int_\Omega f(\omega)\, dm = \lim_{k\to\infty} \int_\Omega s_k(\omega)\, dm = \lim_{k\to\infty} \int_\Omega s_k(\sigma_t(\omega))\, dm = \int_\Omega f(\sigma_t(\omega))\, dm\,.$$

Finally, any function $f \in L^1(\Omega, m)$ can be written as $f = f^+ - f^-$, where $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0)$ are nonnegative elements of $L^1(\Omega, m)$. This proves (2).

(2)⇒(3) This property is obvious.

(3)⇒(1) Property (3) and the Lebesgue monotone convergence theorem yield $m(\mathcal{K}) = m(\sigma_t(\mathcal{K}))$ whenever $\mathcal{K} \subseteq \Omega$ is compact and $t \in \mathbb{R}$: just take a decreasing sequence of positive and continuous functions (f_k) with pointwise limit $\chi_{\mathcal{K}}$ and with bound 1. (For instance, $f_k(\omega) = 1/(1 + k\, d_\Omega(\omega, \mathcal{K}))$, where d_Ω is the distance in Ω and $d_\Omega(\omega, \mathcal{K}) = \inf_{\tilde{\omega}\in\mathcal{K}} d_\Omega(\omega, \tilde{\omega})$.) Hence, the regularity of the measure m ensures the same property for every Borel set $\mathcal{A} \subseteq \Omega$.

The occurrence or lack of invariant and ergodic measures is a fundamental question in measure theory. There are examples of noncontinuous flows on compact metric spaces (see [99], Exercise I.8.6) as well as more basic examples of continuous flows on noncompact spaces which do not admit any normalized invariant measure.

However, the situation is better when dealing with a continuous flow on a compact metric space, as stated in Theorem 1.8. This will be the setting from now on: until the end of this section, (Ω, σ) will represent a continuous flow on a compact metric space. A complete proof of Theorem 1.8 in the case of a discrete flow can be found in [99], Section I.8, and for a real flow in [110], Theorem 9.05 of Chapter VI. In fact the result was initially proved by Krylov and Bogoliubov [94].

Theorem 1.8 *Let σ be a continuous flow on a compact metric space Ω. Then there exists at least one σ-invariant measure on Ω.*

In order to deduce the existence of σ-ergodic measures from the above result, which is one of the assertions of the following theorem, consider the set $\mathfrak{M}(\Omega)$ of positive normalized regular Borel measures on Ω endowed with the weak* topology: the sequence of measures (m_k) converges to m if and only if $\lim_{k\to\infty} \int_\Omega f(\omega)\, dm_k = \int_\Omega f(\omega)\, dm$ for every continuous function $f\colon \Omega \to \mathbb{R}$. Then, $\mathfrak{M}(\Omega)$ is a metrizable compact space (see e.g. Theorems 6.4 and 6.5 of [148]), and it is clearly convex: any convex combination of measures $m_1, \ldots, m_n$ in $\mathfrak{M}(\Omega)$ (i.e. the sum $\lambda_1 m_1 + \ldots \lambda_n m_n$, where $\lambda_1, \ldots, \lambda_n \in [0, 1]$ and $\sum_{j=1}^n \lambda_j = 1$), belongs to $\mathfrak{M}(\Omega)$. Recall that given a convex subset $\mathfrak{M}$ of $\mathfrak{M}(\Omega)$, a point m is *extremal* if the equality $m = a\, m_1 + (1-a)\, m_2$ for $a \in [0, 1]$ and $m_1, m_2 \in \mathfrak{M}$ ensures that $a \in \{0, 1\}$; and that the *closed convex hull* of a subset $\mathfrak{M}_1 \subseteq \mathfrak{M}$ is the closure of the set of convex combinations of points of $\mathfrak{M}_1$.

Theorem 1.9 *Let σ be a continuous flow on a compact metric space Ω.*

(i) *The nonempty set $\mathfrak{M}_{\text{inv}}(\Omega, \sigma)$ of σ-invariant measures is a compact convex subset of $\mathfrak{M}(\Omega)$.*
(ii) *$\mathfrak{M}_{\text{inv}}(\Omega, \sigma)$ is the closed convex hull of the subset of its extremal points.*
(iii) *An element of $\mathfrak{M}_{\text{inv}}(\Omega, \sigma)$ is an extremal point if and only if it is a σ-ergodic measure.*

In particular, there exist σ-ergodic measures, and every σ-invariant measure on Ω can be written as the limit in the weak* topology of a sequence of convex combinations of σ-ergodic measures on Ω.

Proof The proof of points (i) and (iii) can be easily carried out by adapting to the real case the arguments of Theorem 6.10 of [148] for the discrete case. To this end, use Theorem 1.8 and Proposition 1.7. Point (ii) is an immediate consequence of (i) and Krein–Milman theorem (see e.g. Theorem 3.23 of [129]), and the last assertions follow from the previous ones.

Another classical way to deduce the existence of σ-ergodic measures from the existence of σ-invariant ones is to use the Choquet representation theorem.

Remark 1.10 Let σ be a continuous flow on a compact metric space Ω. The *ergodic component* of a σ-invariant measure m on Ω is defined as the set of points $\omega_0 \in \Omega$ such that

$$\int_\Omega f(\omega)\,dm = \lim_{t\to\infty} \frac{1}{t}\int_0^t f(\sigma_s(\omega_0))\,ds$$

for all $f \in C(\Omega,\mathbb{R})$. In other words, it is the intersection of all the σ-invariant sets Ω_f associated by Theorem 1.3 to the continuous functions f. It is not hard to check that Ω_f is a Borel set if f is continuous. The separability of $C(\Omega,\mathbb{R})$ for the topology given by the norm $\|f\|_\Omega = \max_{\omega\in\Omega}|f(\omega)|$ implies that the ergodic component is also a Borel set, and that it has measure 1 in the case that m is ergodic: the ergodicity and Theorem 1.6 ensures that $m(\Omega_f) = 1$ for every continuous function f.

Theorem 1.9 ensures that, if the flow σ is continuous (and Ω is not necessarily compact), any minimal subset $\mathcal{K} \subseteq \Omega$ *concentrates* at least one σ-ergodic measure; that is, there exists a σ-ergodic measure m on Ω such that $m(\mathcal{K}) = 1$. In general, one says that a measure m on Ω *is concentrated on a subset* if this subset has measure 1. Recall that every measure is normalized unless otherwise indicated.

Let m be a measure on a compact metric space. The *topological support* of m, $\operatorname{Supp} m$, is the set $\Omega - \mathcal{O}$, where $\mathcal{O} \subset \Omega$ is the largest open subset with $m(\mathcal{O}) = 0$. Obviously, $\operatorname{Supp} m$ is a compact subset of Ω, with $m(\operatorname{Supp} m) = 1$: the measure is concentrated on its support. In addition,

Proposition 1.11 *Let (Ω,σ) be a continuous flow on a compact metric space and let* $\operatorname{Supp} m$ *be the topological support of a σ-invariant measure m.*

(i) *Suppose that* $\operatorname{Supp} m = \Omega$*, and let $\Omega_0 \subseteq \Omega$ satisfy $m(\Omega_0) = 1$. Then Ω_0 is dense in Ω.*
(ii) *If m is σ-invariant, so is* $\operatorname{Supp} m$*.*
(iii) *If m is σ-invariant and Ω is minimal, then* $\operatorname{Supp} m = \Omega$*. In fact Ω is minimal if and only if any σ-ergodic measure has full support.*

Proof

(i) Suppose for contradiction the existence of a nonempty open subset $\mathcal{O} \subset \Omega$ with $\Omega_0 \cap \mathcal{O}$ empty. Then $m(\mathcal{O}) = 0$, so that $\mathcal{O}$ is contained in the $\Omega - \operatorname{Supp} m$, which is empty. (Note that, in fact, the invariance of the measure is not required for this property.)
(ii) Let $\mathcal{O}$ be as in the definition of $\operatorname{Supp} m$. Then $\sigma_t(\mathcal{O})$ is open and $m(\sigma_t(\mathcal{O})) = m(\mathcal{O}) = 0$ for all $t \in \mathbb{R}$, so that $\sigma_t(\mathcal{O}) = \mathcal{O}$. That is, $\sigma_t(\operatorname{Supp} m) = \operatorname{Supp} m$ for all $t \in \mathbb{R}$.
(iii) $\operatorname{Supp} m$ is compact, since Ω is so. Hence, the first property in (iii) follows from (ii). The "if" assertion follows from Theorems 1.8 and 1.9: if $\mathcal{M} \subsetneq \Omega$ is a compact σ-invariant set, it concentrates a σ-ergodic measure m, and hence $\operatorname{Supp} m \subsetneq \Omega$.

The following property will be required several times in the book. Its proof is included here for the reader's convenience.

Proposition 1.12 *Let (Ω, σ) be a continuous flow on a compact metric space. Suppose that $\Omega = \operatorname{Supp} m$ for a σ-ergodic measure m. Then there exist subsets $\Omega^{\pm} \subseteq \Omega$ with $m(\Omega^{\pm}) = 1$ such that the positive σ-semiorbit of any $\omega \in \Omega^{+}$ and the negative σ-semiorbit of any $\omega \in \Omega^{-}$ are dense in Ω. In particular, (Ω, σ) is chain recurrent. In addition, Ω agrees with the omega-limit of any point $\omega \in \Omega^{+}$ and with the alpha-limit of any point $\omega \in \Omega^{-}$.*

Proof Let $\{\mathcal{O}_k \mid k \geq 1\}$ be a countable basis of open subsets of the compact set Ω. Since $\Omega = \operatorname{Supp} m$, then $m(\mathcal{O}_k) > 0$ for each $k \geq 1$. It follows from the Birkhoff Theorems 1.3 and 1.6 that the set

$$\Omega_k^{+} = \{\omega \in \Omega \mid \sigma_t(\omega) \in \mathcal{O}_k \text{ for some } t > 0\}$$

has measure 1, and hence also the countable intersection $\Omega^{+} = \bigcap_{k \geq 1} \Omega_k$ has full measure for m. Obviously any point in this intersection has dense positive semiorbit. The set Ω^{-} is defined from

$$\Omega_k^{-} = \{\omega \in \Omega \mid \sigma_t(\omega) \in \mathcal{O}_k \text{ for some } t < 0\}\,.$$

The chain recurrence follows easily from the fact that any point in the set $\Omega^{+} \cap \Omega^{-}$ has dense positive and negative semiorbits.

Take now $\omega \in \Omega^{+}$. If its σ-orbit is periodic, then its positive σ-semiorbit is finite and dense, and hence $\Omega = \mathcal{O}(\omega)$. Assume that this is not the case. Then the point $\omega{\cdot}1$ belongs to the closure of the positive semiorbit of ω (which agrees with Ω) but not to the orbit. Therefore $\omega{\cdot}1 \in \mathcal{O}(\omega)$, which ensures that $\{\sigma_t(\omega) \mid t \geq 0\} \subseteq \mathcal{O}(\omega)$. Since $\mathcal{O}(\omega)$ is closed, it follows that $\Omega = \operatorname{closure}_{\Omega}\{\sigma_t(\omega) \mid t \geq 0\} \subseteq \mathcal{O}(\omega)$. This proves the last assertion in the case of Ω^{+}, and a similar argument proves it in the case of Ω^{-}.

Remarks 1.13

1. Note that in fact the last argument of the previous proof shows that if the positive (resp. negative) σ-semiorbit of a point $\omega \in \Omega$ is dense, then $\mathcal{O}(\omega) = \Omega$ (resp. $\mathcal{A}(\omega) = \Omega$).
2. It is easy to check that if Ω reduces to a point or is composed of just one periodic σ-orbit, then it admits a unique σ-invariant measure, which therefore is ergodic. In addition, it turns out that it has full topological support.

In most of the sections of the book, (Ω, σ) will indicate a fixed continuous flow on a compact metric space. The representation

$$\omega{\cdot}t = \sigma_t(\omega)$$

will be used from now on when no confusion may arise.

1.1.3 Skew-Product Flows

Let Ω and $\mathbb{Y}$ satisfy the conditions imposed on Ω in the previous section: they are locally compact Hausdorff topological spaces which admit countable bases of open sets. Hence, $\Omega \times \mathbb{Y}$ satisfies the same properties. In what follows, the product space $\Omega \times \mathbb{Y}$ is understood as a bundle over Ω: this is done throughout the book for several different spaces $\mathbb{Y}$. The sets Ω and $\mathbb{Y}$ will be referred to respectively as the *base* and the *fiber* of the bundle.

Let σ be a Borel measurable flow on Ω. A *skew-product flow on* $\Omega \times \mathbb{Y}$ *projecting onto* σ is a Borel measurable real flow

$$\tilde{\tau}: \mathbb{R} \times \Omega \times \mathbb{Y} \to \Omega \times \mathbb{Y}, \quad (\omega, y) \mapsto (\omega{\cdot}t, \tilde{\tau}_2(t, \omega, y)) .$$

The flow (Ω, σ) is the *base flow* of $\tilde{\tau}$. Note that $\tilde{\tau}_2$ satisfies $\tilde{\tau}_2(s + t, \omega, y) = \tilde{\tau}_2(s, \omega{\cdot}t, \tau(t, \omega, y))$.

Some results concerning noncontinuous skew-product flows will be required in Chap. 4, and explained in the appropriate place. For the time being, let Ω and $\mathbb{Y}$ be compact metric spaces, and let $\tilde{\tau}$ be a continuous skew-product flow on $\Omega \times \mathbb{Y}$ with continuous base flow (Ω, σ). It is easy to check that, given a measure μ on $\Omega \times \mathbb{Y}$, the relation $m(\mathcal{A}) = \mu(\mathcal{A} \times \mathbb{Y})$ for every Borel set $\mathcal{A} \subseteq \Omega$ defines a measure m on Ω, which in addition is σ-invariant if μ is $\tilde{\tau}$-invariant. In this case, it is said that μ *projects onto* m.

Remark 1.14 In fact, μ projects onto m if and only if $\int_{\Omega\times\mathbb{Y}} f(\omega)\, d\mu = \int_{\Omega} f(\omega)\, dm$ for all $f \in C(\Omega, \mathbb{R})$. For the "if" assertion, keep in mind the regularity of the measures (see the proof of (3)⇒(1) in Proposition 1.7). The "only if" assertion is an easy consequence of the Lebesgue monotone convergence theorem (see the proof of (1)⇒(2) in Proposition 1.7).

The following result, whose proof is included for the reader's convenience, presents the well-known construction of a $\tilde{\tau}$-invariant measure projecting onto a fixed σ-ergodic measure m on Ω.

Proposition 1.15 *Let Ω and $\mathbb{Y}$ be compact metric spaces, and let $\tilde{\tau}$ be a continuous skew-product flow on $\Omega \times \mathbb{Y}$ with continuous base flow (Ω, σ). Let m be a σ-ergodic measure on Ω. Then,*

(i) *there exist $\tilde{\tau}$-invariant measures on $\Omega \times \mathbb{Y}$ projecting onto m.*
(ii) *The set $\mathfrak{M}_{\mathrm{inv},m}(\Omega \times \mathbb{Y}, \tilde{\tau})$ of the $\tilde{\tau}$- invariant measures projecting onto m is a convex compact set in the weak* topology.*
(iii) *There exist σ-ergodic measures on $\Omega \times \mathbb{Y}$ projecting onto m, and every $\tilde{\tau}$-invariant measure projecting onto m can be written as the limit in the weak* topology of a sequence of convex combinations of $\tilde{\tau}$-ergodic measures projecting onto m.*

Proof

(i) The Birkhoff Theorems 1.3 and 1.6 and the ergodicity of the measure m ensure that the set

$$\Omega_c = \left\{ \omega_0 \in \Omega \;\middle|\; \lim_{t\to\infty} \frac{1}{2t} \int_{-t}^{t} f(\omega_0{\cdot}s)\, ds = \int_\Omega f(\omega)\, dm \quad \forall f \in C(\Omega, \mathbb{R}) \right\}.$$

is σ-invariant and that $m(\Omega_c) = 1$: see Remark 1.10. Now fix $(\omega_0, y_0) \in \Omega_c \times \mathbb{Y}$. Let $C(\Omega \times \mathbb{Y}, \mathbb{R})$ be the set of real continuous functions on the space $\Omega \times \mathbb{Y}$. Take also a sequence $(t_k) \uparrow \infty$. The Riesz representation theorem associates to the bounded linear functional defined by $C(\Omega\times\mathbb{Y}, \mathbb{R}) \to \mathbb{R}$, $\tilde f \mapsto (1/(2t_k)) \int_{-t_k}^{t_k} \tilde f(\tilde\tau(s, \omega_0, y_0))\, ds$, whose norm is 1, a (positive normalized regular Borel) measure μ_k, which satisfies

$$\int_{\Omega\times\mathbb{Y}} \tilde f(\omega, y)\, d\mu_k = \frac{1}{2t_k} \int_{-t_k}^{t_k} \tilde f(\tilde\tau(s, \omega_0, y_0))\, ds\,.$$

As stated in the previous section, the set of (positive normalized regular Borel) measures on $\Omega \times \mathbb{Y}$ is a metrizable compact set in the weak* topology. Therefore, the sequence (μ_k) admits a subsequence (μ_j) which converges weak* to a measure μ. That is,

$$\int_{\Omega\times\mathbb{Y}} \tilde f(\omega, y)\, d\mu = \lim_{j\to\infty} \frac{1}{2t_j} \int_{-t_j}^{t_j} \tilde f(\tilde\tau(s, \omega_0, y_0))\, ds$$

whenever $\tilde f \in C(\Omega \times \mathbb{Y}, \mathbb{R})$. It follows easily from this fact, from the condition $\omega_0 \in \Omega_c$, and from Remark 1.14, that μ projects onto m. Note also that, if $l \in \mathbb{R}$ and $\tilde f \in C(\Omega \times \mathbb{Y}, \mathbb{R})$, then

$$\begin{aligned}\int_{\Omega\times\mathbb{Y}} \tilde f \circ \tilde\tau_l(\omega, y)\, d\mu &= \lim_{j\to\infty} \frac{1}{2t_j} \int_{-t_j}^{t_j} \tilde f(\tilde\tau(s + l, \omega_0, y_0))\, ds \\ &= \lim_{j\to\infty} \frac{1}{2t_j} \int_{-t_j-l}^{t_j-l} \tilde f(\tilde\tau(s, \omega_0, y_0))\, ds = \int_{\Omega\times\mathbb{Y}} \tilde f(\omega, y)\, d\mu\,,\end{aligned}$$

as can be deduced from the boundedness of $\tilde f$. According to Proposition 1.7, this equality proves the $\tilde\tau$-invariance of μ, and completes the proof of (i).

(ii) Let $\mathfrak{M}_{\text{inv},m}(\Omega \times \mathbb{Y}, \tilde\tau)$ be the set of the $\tilde\tau$-invariant measures on $\Omega \times \mathbb{Y}$ which project onto m. As in Theorem 1.9(i), an immediate application of the implication (3)⇒(1) of Proposition 1.7 proves that if a measure μ is the limit in the weak* topology of a sequence (μ_k) of elements of $\mathfrak{M}_{\text{inv},m}(\Omega \times \mathbb{Y}, \tilde\tau)$, then $\mu \in \mathfrak{M}_{\text{inv}}(\Omega \times \mathbb{Y}, \tilde\tau)$. In order to check that μ projects onto m, keep in mind Remark 1.14, and note that if $f \in C(\Omega, \mathbb{R})$, then $\int_\Omega f(\omega)\, d\mu =$

$\lim_{k\to\infty}\int_\Omega f(\omega)\,d\mu_k = \lim_{k\to\infty}\int_\Omega f(\omega)\,dm = \int_\Omega f(\omega)\,dm$. The convexity of $\mathfrak{M}_{\mathrm{inv},m}(\Omega\times\mathbb{Y},\tilde\tau)$ is clear.

(iii) As in the proof of Theorem 1.9(ii), properties (i) and (ii) allow one to apply the Krein–Milman theorem to prove that the nonempty convex compact set $\mathfrak{M}_{\mathrm{inv},m}(\Omega\times\mathbb{Y},\tilde\tau)$ agrees with the closed convex hull of the subset of its extremal points. In addition, these extremal points are precisely the σ-ergodic measures projecting onto m. To prove this, assume first that the measure $\mu\in\mathfrak{M}_{\mathrm{inv},m}(\Omega\times\mathbb{Y},\tilde\tau)$ is ergodic, and apply Theorem 1.9(iii) to deduce that it is extremal in $\mu\in\mathfrak{M}_{\mathrm{inv}}(\Omega\times\mathbb{Y},\tilde\tau)$, which obviously ensures that it is extremal in $\mathfrak{M}_{\mathrm{inv},m}(\Omega\times\mathbb{Y},\tilde\tau)$. Conversely, assume that a measure μ is extremal in $\mathfrak{M}_{\mathrm{inv},m}(\Omega\times\mathbb{Y},\tilde\tau)$ and that $\mu = a\,\mu_1 + (1-a)\,\mu_2$ for $a\in[0,1]$ and $\mu_1,\,\mu_2\in\mu\in\mathfrak{M}_{\mathrm{inv}}(\Omega\times\mathbb{Y},\tilde\tau)$. Then $m = a\,m_1 + (1-a)\,m_2$, where m_1 and m_2 are the σ-invariant measures on Ω defined by the projections of μ_1 and μ_2. Since m is σ-ergodic, Theorem 1.9(iii) ensures that $a\in\{0,1\}$, so that μ is extremal in $\mathfrak{M}_{\mathrm{inv}}(\Omega\times\mathbb{Y},\tilde\tau)$ and hence $\tilde\tau$-ergodic. The assertions in (iii) are proved.

Proposition 1.16 *Let Ω and $\mathbb{Y}$ be compact metric spaces, and let $\tilde\tau$ be a continuous skew-product flow on $\Omega\times\mathbb{Y}$ with continuous base flow (Ω,σ). Let m be a σ-ergodic measure on Ω. Let $\Omega_0\in\Sigma_m$ be a σ-invariant set with $m(\Omega_0)=1$, and let $l\colon\Omega\to\mathbb{Y}$ be a Σ_m-measurable map with $\tilde\tau(t,\omega,l(\omega))=(\omega{\cdot}t,l(\omega{\cdot}t))$ for all $\omega\in\Omega_0$. Then,*

(i) *there exists a Borel set $\Omega_1\subseteq\Omega_0$ which is σ-invariant set and with $m(\Omega_1)=1$ such that the $\tilde\tau$-invariant set $\{(\omega,l(\omega))\mid\omega\in\Omega_1\}\subset\Omega\times\mathbb{Y}$ is Borel.*

(ii) *The graph of l concentrates a $\tilde\tau$-invariant measure μ_l which projects onto m, which is determined by $\int_{\Omega\times\mathbb{Y}}\tilde f(\omega,y)\,d\mu_l=\int_\Omega\tilde f(\omega,l(\omega))\,dm$ for all continuous functions $\tilde f\colon\Omega\times\mathbb{Y}\to\mathbb{R}$.*

Proof

(i) The regularity of m and Lusin's theorem guarantee the existence of a compact subset $\mathcal{M}\subseteq\Omega_0$ with $m(\mathcal{M})>0$ such that l is continuous at the points of $\mathcal{M}$. Define $\mathcal{M}_k=\{\omega{\cdot}t\mid\omega\in\mathcal{M},t\in[-k,k]\}$ for $k=0,1,2,\ldots$, which is also a compact set, since σ is continuous. It is easy to check that $\Omega_1=\cup_{k\ge0}\mathcal{M}_k$ is a Borel σ-invariant set of positive measure; hence, by ergodicity, $m(\Omega_1)=1$. In addition, the map l is continuous at the points of all the sets $\mathcal{M}_k$, as one can easily deduce from the property of τ-invariance of l on Ω_0 and from the compactness of $\mathcal{M}$ and $[-k,k]$. Finally, one has that $\{(\omega,l(\omega))\mid\omega\in\Omega_1\}=\cup_{k\ge0}\{(\omega,l(\omega))\mid\omega\in\mathcal{M}_k\}$, and therefore it is Borel.

(ii) A $\tilde\tau$-invariant measure concentrated on $\mathcal{L}=\{(\omega,l(\omega))\mid\omega\in\Omega_1\}$, which is contained in the graph of l, is constructed in what follows. For all $\tilde f$ in the set $C(\Omega\times\mathbb{Y},\mathbb{R})$ of real continuous functions, the Σ_m-measurable map $\Omega\to\mathbb{R},\ \omega\mapsto\tilde f(\omega,l(\omega))$ belongs to $L^1(\Omega,\mathbb{R})$, so that it is possible to define $L(\tilde f)=\int_\Omega\tilde f(\omega,l(\omega))\,dm$. Then L is a bounded linear functional with norm 1 on $C(\Omega\times\mathbb{Y},\mathbb{R})$, and the Riesz representation theorem provides a (positive normalized regular Borel) measure μ_l such that $L(\tilde f)=\int_{\Omega\times\mathbb{Y}}\tilde f(\omega,y)\,d\mu_l$.

Since $\int_{\Omega\times\mathbb{Y}} f(\omega)\,d\mu_l = L(f) = \int_\Omega f(\omega)\,dm$ for all $f \in C(\Omega,\mathbb{R})$, Remark 1.14 ensures that μ_l projects onto m. In addition, according to the equivalences established in Proposition 1.7, the measure μ_l is $\tilde{\tau}$-invariant: if $\tilde{f} \in C(\Omega\times\mathbb{Y},\mathbb{R})$ and $t \in \mathbb{R}$, then $\int_{\Omega\times\mathbb{Y}} \tilde{f}(\tilde{\tau}_t(\omega,y))\,d\mu_l = \int_\Omega \tilde{f}(\omega{\cdot}t, l(\omega{\cdot}t))\,dm = \int_\Omega \tilde{f}(\omega, l(\omega))\,dm = \int_{\Omega\times\mathbb{Y}} \tilde{f}(\omega,y)\,d\mu_l$, since m is σ-invariant. Finally, if $\mathcal{K} \subseteq (\Omega\times\mathbb{Y}) - \mathcal{L}$ is a compact set, the Lebesgue monotone convergence theorem ensures that $\mu_l(\mathcal{K}) = 0$ (see the proof of (3)$\Rightarrow$ (1) in Proposition 1.7), and hence the regularity of μ_l ensures that $\mu_l(\mathcal{L}) = 1$; that is, μ_l is concentrated on $\mathcal{L}$.

A skew-product flow may admit many types of compact invariant sets, whose complexity varies in an ample range. Among them are those described now, which are especially interesting from a dynamical point of view, and which will appear frequently in the following chapters. The first one represents an extension of the idea of an equilibrium point for an autonomous system, or of a T-periodic solution for a system with T-periodic coefficients.

Definition 1.17 Let Ω and $\mathbb{Y}$ be compact metric spaces, and let $\tilde{\tau}$ be a continuous skew-product flow on $\Omega\times\mathbb{Y}$ with continuous base flow (Ω,σ). A compact subset $\mathcal{K} \subset \Omega\times\mathbb{Y}$ is a *copy of the base (for the flow $\tilde{\tau}$)* if it is $\tilde{\tau}$-invariant and, in addition, $\mathcal{K}_\omega = \{y \in \mathbb{Y} \mid (\omega,y) \in \mathcal{K}\}$ reduces to a point for every $\omega \in \Omega$: in other words, if it agrees with the graph of a continuous map $c\colon \Omega \to \mathbb{Y}$ satisfying $c(\omega{\cdot}t) = \tilde{\tau}_2(t,\omega,c(\omega))$, so that $\mathcal{K} = \{(\omega,c(\omega)) \mid \omega \in \Omega\}$.

These invariant objects are the simplest ones from a dynamical point of view, since they reproduce homeomorphically the base Ω. The second type of set is a generalization of the first one.

Definition 1.18 Let Ω and $\mathbb{Y}$ be compact metric spaces, and let $\tilde{\tau}$ be a continuous skew-product flow on $\Omega\times\mathbb{Y}$ with continuous and minimal base flow (Ω,σ). A minimal subset $\mathcal{K} \subset \Omega\times\mathbb{Y}$ is an *almost automorphic extension of the base (for the flow $\tilde{\tau}$)* if it is $\tilde{\tau}$-invariant and, in addition, there exists $y \in \mathbb{Y}$ such that $\mathcal{K}_\omega = \{y \in \mathbb{Y} \mid (\omega,y) \in \mathcal{K}\}$ reduces to a point.

Clearly, a copy of the base provides the simplest example of an almost automorphic extension in the case of minimal base flow. However, there are examples of almost automorphic extensions which are not copies of the base. The most classical ones are those due to Millionščikov [104, 105] and Vinograd [147]. See Johnson [68] for a detailed dynamical description of these examples, and [67] for a later example of a scalar linear equation with this type of complicated invariant object, which can exhibit properties of high dynamical complexity (like sensitive dependence with respect to initial conditions). Example 8.44 contains a similar construction with most of the details explained, with an almost automorphic extension of the base whose fibers reduce to a singleton at a residual set of points of the base but not on a set of full measure.

1.2 Basic Properties of Matrices and Lagrange Planes

Throughout the book, $\mathbb{M}_{m\times d}(\mathbb{R})$ and $\mathbb{M}_{m\times d}(\mathbb{C})$ will represent the $(m \times d)$-dimensional vector spaces of real or complex $m \times d$ matrices; and, as in the previous pages, the symbol $\mathbb{K}$ will represent either $\mathbb{R}$ or $\mathbb{C}$. The cases $d = m = n$ and $d = m = 2n$ will be frequently considered. The symbols M^T and M^* denote respectively the transpose and conjugate transpose of M; and $\operatorname{Re} M$ and $\operatorname{Im} M$ represent the real and imaginary parts of a complex matrix M. The determinant and trace of a square matrix M will be represented by $\det M$ and $\operatorname{tr} M$. Recall that $\det M = \det M^T$, $\det(MN) = \det(NM)$, $\operatorname{tr} M = \operatorname{tr} M^T$, and $\operatorname{tr}(MN) = \operatorname{tr}(NM)$. If M is a *nonsingular* square matrix (i.e. if $\det M \neq 0$), then M^{-1} represents its inverse; I_d and 0_d are the identity and null matrices in $\mathbb{M}_{d\times d}(\mathbb{K})$ for all $d \in \mathbb{N}$; and $\mathbf{0}$ represents the null vector in $\mathbb{K}^d$ for all $d \in \mathbb{N}$.

A $d \times d$ matrix M is *symmetric* if $M^T = M$ and *hermitian* if $M^* = M$. A real symmetric matrix or a complex hermitian matrix is *selfadjoint*, in reference to the usual Euclidean inner product $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{y}^*\mathbf{x}$ in $\mathbb{K}^d$: in both cases $\langle \mathbf{x}, M\mathbf{y}\rangle = \langle M\mathbf{x}, \mathbf{y}\rangle$ for any pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{K}^d$. The square matrix M is *unitary* when $M^*M = I_d$, and *orthogonal* if it is real and unitary. A real or complex $2n \times 2n$ matrix M is *symplectic* if $M^TJM = J$, where

$$J = \begin{bmatrix} 0_n & -I_n \\ I_n & 0_n \end{bmatrix}.$$

Note that $J^2 = -I_{2n}$, and that any symplectic matrix is nonsingular: in fact, $\det M = 1$. This can be deduced, for instance, from the Iwasawa decomposition of M, described in Lemma 2.16. The simplest examples of symplectic matrices are I_{2n} and J.

The following notation will always be used:

- $\mathrm{GL}(m, \mathbb{K})$: set of nonsingular $m \times m$ matrices con coefficients in $\mathbb{K}$,
- $\mathrm{U}(m, \mathbb{C})$: set of (complex) unitary $m \times m$ matrices,
- $\mathrm{SU}(m, \mathbb{C})$: set of unitary $m \times m$ matrices with determinant 1,
- $\mathrm{O}(m, \mathbb{R})$: set of (real) orthogonal $m \times m$ matrices,
- $\mathrm{SO}(m, \mathbb{R})$: set of orthogonal $m \times m$ matrices with determinant 1,
- $\mathrm{Sp}(n, \mathbb{C})$: set of complex symplectic $2n \times 2n$ matrices,
- $\mathrm{Sp}(n, \mathbb{R})$: set of real symplectic $2n \times 2n$ matrices.

It is very easy to check that the first five sets are groups with respect to the matrix product. Proposition 1.23 below guarantees that also $\mathrm{Sp}(n, \mathbb{C})$ and $\mathrm{Sp}(n, \mathbb{R})$ are groups. In fact, all of them are Lie groups (see e.g. Sections 1 and 2 of Chapter II of Helgason [57]).

1.2.1 Symmetric, Hermitian, and Symplectic Matrices

Consider the sets of real and complex symmetric $d \times d$ matrices,

$$\mathbb{S}_d(\mathbb{R}) = \{M \in \mathbb{M}_{d\times d}(\mathbb{R}) \mid M = M^T\},$$
$$\mathbb{S}_d(\mathbb{C}) = \{M \in \mathbb{M}_{d\times d}(\mathbb{C}) \mid M = M^T\},$$

which constitute $(d \times (d+1))/2$-dimensional linear subspaces of $\mathbb{M}_{d\times d}(\mathbb{R})$ and $\mathbb{M}_{d\times d}(\mathbb{C})$. Let M belong to $\mathbb{S}_d(\mathbb{R})$. Then,

- M is a *positive definite matrix* ($M > 0$) if $\mathbf{x}^T M\mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{x} \neq \mathbf{0}$,
- M is a *positive semidefinite matrix* ($M \geq 0$) if $\mathbf{x}^T M\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^d$,
- M is a *negative definite matrix* ($M < 0$) if $-M > 0$,
- M is a *negative semidefinite matrix* ($M \leq 0$) if $-M \geq 0$.

The subsets

$$\mathbb{S}_d^+(\mathbb{R}) = \{M \in \mathbb{S}_d(\mathbb{R}) \mid M > 0\},$$
$$\mathbb{S}_d^+(\mathbb{C}) = \{M \in \mathbb{S}_d(\mathbb{C}) \mid \operatorname{Im} M > 0\}$$

will be frequently considered. Note that their closures on $\mathbb{S}_d(\mathbb{R})$ and $\mathbb{S}_d(\mathbb{C})$ are given by

$$\overline{\mathbb{S}_d^+}(\mathbb{R}) = \{M \in \mathbb{S}_d(\mathbb{R}) \mid M \geq 0\},$$
$$\overline{\mathbb{S}_d^+}(\mathbb{C}) = \{M \in \mathbb{S}_d(\mathbb{C}) \mid \operatorname{Im} M \geq 0\},$$

which can be easily deduced from the definitions of positive definite and semidefinite matrices.

Similarly, if M is a hermitian matrix, then

- M is *positive definite* (($M > 0$) if $\mathbf{x}^* M\mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{C}^d$, $\mathbf{x} \neq \mathbf{0}$,
- M is *positive semidefinite* (($M \geq 0$) if $\mathbf{x}^* M\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{C}^d$,
- M is *negative definite* (($M < 0$) if $-M > 0$,
- M is *negative semidefinite* (($M \geq 0$) if $-M \geq 0$.

The relations $M > N$, $M \geq N$, $M < N$ and $M \leq N$ for selfadjoint matrices have the obvious meaning: for instance, $M > N$ means that $M - N > 0$.

Well-known properties of positive (real or complex) matrices are: $M > 0$ (resp. $M \geq 0$) if and only if $\lambda > 0$ (resp. $\lambda \geq 0$) for all the eigenvalues λ of M (which are real); hence, $M > I_d$ (resp. $M \geq I_d$) if and only if $\lambda > 1$ (resp. $\lambda \geq 1$) for all the eigenvalues λ of M; and, if $M > 0$ (resp. $M \geq 0$) and P is nonsingular, then $P^* MP > 0$ (resp. $P^* MP \geq 0$).

Let $\mathbb{M}_{d\times d}(\mathbb{K})$ be a given positive definite (or semidefinite) matrix. Throughout the book, the expression "the unique positive definite (or semidefinite) square root of

M" (will be used very often. That this object exists for positive selfadjoint bounded operators is a well-known fact: see e.g. Theorem VI.9 of Reed and Simon [122]. The following result provides an easy and constructive proof in the matrix case.

Proposition 1.19 *Let M belong to $\mathbb{M}_{d\times d}(\mathbb{C})$.*

(i) *If $M \geq 0$, there exists a unique matrix $M^{1/2} \geq 0$ such that $(M^{1/2})^2 = M$. In addition, $M\mathbf{x} = \mathbf{0}$ if and only if $M^{1/2}\mathbf{x} = \mathbf{0}$, and $M^{1/2}$ is real if M is real.*
(ii) *If $M > 0$, then $M^{1/2} > 0$ and $(M^{1/2})^{-1} = (M^{-1})^{1/2}$.*
(iii) *The map defined from $\{M \in \mathbb{S}_d(\mathbb{C}) \mid M > 0\}$ to itself by sending M to $M^{1/2}$ is continuously differentiable.*

Proof If $M \geq 0$, all its eigenvalues are real and nonnegative, and it is a well-known fact that there exist a unitary matrix P (which is real if M is real) and a real diagonal matrix $D \geq 0$ such that $M = P^*DP$. It is obvious that there exists a unique diagonal matrix $\widetilde{D} \geq 0$ such that $D = \widetilde{D}^2$. Then the matrix $N = P^*\widetilde{D}P$ satisfies $N \geq 0$ and $N^2 = M$. This proves the existence, and the fact that N is real if M is real. Clearly, if $M > 0$, then $D > 0$ and hence $N > 0$. Now, if $N \geq 0$ satisfies $N^2 = M$, and $N\mathbf{x} = \lambda\,\mathbf{x}$, then $M\mathbf{x} = \lambda^2\,\mathbf{x}$. That is, the eigenvalues and the associated eigenvectors of N are uniquely determined, so that also the matrix is. In addition, if $M = P^*\widetilde{D}^2P > 0$, then $M^{-1} = P^*\widetilde{D}^{-2}P > 0$, so that $(M^{1/2})^{-1} = P^*\widetilde{D}^{-1}P = (M^{-1})^{1/2}$. Note also that $M\mathbf{x} = 0$ ensures that $\|M^{1/2}\mathbf{x}\| = 0$ for the Euclidean norm in $\mathbb{C}^d$, so that $M^{1/2}\mathbf{x} = \mathbf{0}$. These facts prove (i) and (ii). To prove (iii), one can apply the Inverse Function Theorem to the map $\mathbb{M}_{d\times d}(\mathbb{C}) \to \mathbb{M}_{d\times d}(\mathbb{C}),\ M \mapsto M^2$ at a point $M > 0$: it is continuously differentiable at M, and its differential, which sends $C \in \mathbb{M}_{d\times d}(\mathbb{C})$ to $MC + CM$, has no null eigenvalues. This last assertion is due to the positive definite character of M: assume that $MC + CM = 0_d$ in order to deduce that $DPCP^* + PCP^*D = 0_d$; and note that this implies that $PCP^* = 0_d$ and hence that $C = 0_d$.

Remark 1.20 Suppose that $0 < M \leq N$ for two symmetric $d \times d$ matrix-valued functions. Then $I_d \leq M^{-1/2}NM^{-1/2}$ and hence $I_d \leq N^{1/2}M^{-1}N^{1/2}$, since both right-hand terms have the same eigenvalues. Therefore, $0 < N^{-1} \leq M^{-1}$. Clearly, there is an analogous result if the inequality is strict.

Proposition 1.21

(i) *If $M \in \mathbb{S}_d^+(\mathbb{C})$, then it is nonsingular, and $-M^{-1} \in \mathbb{S}_d^+(\mathbb{C})$.*
(ii) *If $M \in \overline{\mathbb{S}_d^+}(\mathbb{C})$ is nonsingular, then $-M^{-1} \in \overline{\mathbb{S}_d^+}(\mathbb{C})$.*

Proof Write $M = A + iB$ for real symmetric matrices A and B, and assume that $B > 0$. Take $\mathbf{z} \in \mathbb{C}^n$ with $M\mathbf{z} = \mathbf{0}$ and note that $\mathbf{z}^*M^* = \mathbf{0}^*$. Then $2iB = M - \overline{M} = M - M^*$, and therefore $2i\,\mathbf{z}^*B\,\mathbf{z} = \mathbf{z}^*(M - M^*)\,\mathbf{z} = 0$, so that $\mathbf{z} = \mathbf{0}$. This proves the existence of M^{-1}. To check the second assertion in (i), as well as (ii), write $M^{-1} = C + iD$ for real matrices C and D. It follows from the identity $I_d = (A + iB)(C + iD)$ that $AC - BD = I_d$ and $BC + AD = 0_d$, so that also $C^TB + D^TA = 0_d$. These equalities ensure that $D^T + D^TBD = D^TAC = -C^TBC$, so that $D^T = -D^TBD - C^TBC$, which is obviously symmetric, and is negative semidefinite if $B \geq 0$. Finally, if $B > 0$ and

$\mathbf{z}^* D\,\mathbf{z} = 0$, then $C\mathbf{z} = \mathbf{0}$ and $D\mathbf{z} = \mathbf{0}$, so that $M^{-1}\,\mathbf{z} = \mathbf{0}$ and hence $\mathbf{z} = \mathbf{0}$; that is, $D < 0$.

The following basic properties refer to symplectic matrices.

Proposition 1.22 *If λ is an eigenvalue of $M \in \mathrm{Sp}(n, \mathbb{R})$, so is λ^{-1}.*

Proof Recall that any eigenvalue is different from zero. It can immediately be checked that $M^T J\,\mathbf{v} = \lambda^{-1} J\mathbf{v}$ if $M\,\mathbf{v} = \lambda\,\mathbf{v}$, so that the assertion follows from the coincidence of the set of eigenvalues of any matrix and that of its transpose.

Proposition 1.23 *Let $M = \left[\begin{smallmatrix} M_1 & M_3 \\ M_2 & M_4 \end{smallmatrix}\right]$ belong to $\mathrm{Sp}(n, \mathbb{C})$. Then M^T, M^* and M^{-1} are also symplectic matrices, and*

$$
\begin{aligned}
&M_1^T M_2 = M_2^T M_1 \;\; M_3^T M_4 = M_4^T M_3 \;\; M_1^T M_4 - M_2^T M_3 = I_n \\
&M_1 M_3^T = M_3 M_1^T \;\; M_2 M_4^T = M_4 M_2^T \;\; M_4 M_1^T - M_2 M_3^T = I_n \\
&M_4 M_2^T = M_2 M_4^T \;\; M_3 M_1^T = M_1 M_3^T \\
&M_4^T M_3 = M_3^T M_4 \;\; M_2^T M_1 = M_1^T M_2
\end{aligned}
$$

Proof If $M^T J M = J$, then $J M^T J = -M^{-1}$. This implies, on the one hand, that $M J M^T = J$ and $\overline{M} J M^* = J$, so that M^T and M^* are symplectic whenever M is. And, on the other hand, that $M^{-1} J (M^{-1})^T = J$, so that also $(M^{-1})^T$, and hence M^{-1}, are symplectic if M is.

One more consequence of the identity $M^{-1} = -J M^T J$ is that $M^{-1} = \left[\begin{smallmatrix} M_4^T & -M_3^T \\ -M_2^T & M_1^T \end{smallmatrix}\right]$. The remaining equalities are immediate consequences of the symplectic character of M, M^T, M^{-1}, and $(M^{-1})^T$.

Remarks 1.24

1. Unless otherwise indicated, $\|\cdot\| = \|\cdot\|_d$ will denote throughout the book some fixed norm on the vector space $\mathbb{K}^d$, and $\mathbb{M}_{m\times d}(\mathbb{K})$ will be provided with the associated operator norm, defined by $\|M\| = \max_{\|\mathbf{x}\|_d = 1} \|M\,\mathbf{x}\|_m$. In general, no reference to the dimension will be made in the norm notation: the context will give the precise dimension d or $m \times d$. It can immediately be checked that, with this definition, $\|M\mathbf{x}\| \le \|M\|\,\|\mathbf{x}\|$, and hence that $\|MN\| \le \|M\|\,\|N\|$. Recall that all the norms are equivalent in the case of vector spaces of finite dimension. However, not every norm on $\mathbb{M}_{m\times d}(\mathbb{K})$ is associated as above to a vector norm.
2. The most frequently used norm will be the *Euclidean norm*, defined by $\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x}\rangle^{1/2} = (\mathbf{x}^*\mathbf{x})^{1/2}$ on $\mathbb{K}^d$ and by $\|M\| = \max_{\|\mathbf{x}\|=1} \|M\,\mathbf{x}\|$ on $\mathbb{M}_{d\times m}(\mathbb{K})$. It is the norm associated to the Euclidean inner product defined on $\mathbb{K}^d$ by $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{y}^*\mathbf{x}$. In this case, $\|M\| = \|M^*\|$: according to the Cauchy–Schwarz inequality, $\|M\mathbf{x}\|^2 = \langle \mathbf{x}, M^* M\,\mathbf{x}\rangle \le \|\mathbf{x}\|\,\|M^* M\,\mathbf{x}\| \le \|\mathbf{x}\|\,\|M^*\|\,\|M\,\mathbf{x}\|$, so that $\|M\mathbf{x}\| \le \|\mathbf{x}\|\,\|M^*\|$, which implies $\|M\| \le \|M^*\|$; and hence also $\|M^*\| \le \|(M^*)^*\| = \|M\|$. It is a well-known result that, if M is a square matrix, then $\|M\|^2$ agrees with the spectral radius $\rho(M^* M)$ of $M^* M$; i.e. with the maximum eigenvalue of the matrix