UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ON THE EXISTENCE OF KAM TORI
FOR PRESYMPLECTIC VECTOR FIELDS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

SEAN MICHAEL BAUER
Norman, Oklahoma
2016

ON THE EXISTENCE OF KAM TORI
FOR PRESYMPLECTIC VECTOR FIELDS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF MATHEMATICS

BY

Dr. Nikola P. Petrov, Chair

Dr. Kevin Grasse

Dr. Alex Grigo

Dr. S. Lakshmivarahan

Dr. Christian Remling

Dedicated to Jenny. Love.

Acknowledgements

Graduate school has comprised so much of my adult life, that it is both surreal and a relief to be finally finished. Along the way I have had the pleasure to get to know many great people, learn some mathematics, explore the world, and receive the love and support of my friends and family. I would like to take the time to thank a few specific people, but to also acknowledge the many people that helped me in some way up until this point. I did not take the shortest possible path to get where I am at now, but the scenery has been beautiful.

First, I would like to thank my advisor, Nikola Petrov, without whose help this would not be possible. Nikola went above and beyond the normal role of an advisor, meeting with me over weekends and in the evenings at my favorite coffee shop (C&J) to make sure I finish. Over the last several years we have discussed everything from math to physics to politics to music to life. Thank you for your help and guidance. I can honestly say that I started out with an advisor and ended up with a life-long friend.

I would also like to thank the members of my committee. I have learned many interesting and useful pieces of math from each of you.

To Mom, thank you for being my biggest fan and always making sure that I'm not hungry.

To Umma, thank you for cooking for us and helping to watch Gemma so that

I could study.

To Appa, thank you for all of your support and interest in the topics that I was studying.

To DD, thank you for being proud of me whether I accomplished anything or not.

To Eric, thank you for buying me a bike and teaching me how to barbecue.

To Bailey, thank you for taking me out to eat at the Caf all of those times.

To John Paul, thank you for including me in your dissertation acknowledgments. I hope I can return the favor some day.

To Bubba, thank you for being the world's sweetest baby and for sleeping so well these past several months allowing me to focus on finishing this manuscript.

And last, but most importantly, to Jenny, thank you for everything. Throughout grad school you have been my biggest support in all respects: emotionally, spiritually, physically (you must have woken me up hundreds of times to remind me to fight through the tiredness and keep after it. I guess you are still trying to make up for that Econ test in 2005...). You have bore my burdens and shared in my joy just as much, if not more, than I have. Thank you for being right by my side every step of the way. To you, I dedicate this dissertation.

Abstract

We prove the existence of a torus that is invariant with respect to the flow of a presymplectic vector field found in a family of presymplectic vector fields. Moreover, the flow on this invariant torus is conjugate to a linear flow on a torus with a Diophantine velocity vector. This torus is constructed by iteratively solving functional equations using a Newton method in a space of functions by starting from a torus that is approximately invariant. In contrast to the classical methods of proof, this method does not assume that the system is close to integrable and does not rely on using action-angle variables. The geometry of the problem is used to simplify the equations that come from the Newton method. This method of proof can be implemented into efficient numerical algorithms.

Contents

Chapter 1

Introduction

1.1 KAM theory

1.1.1 A brief history of KAM theory

Near the end of the 19th century, Henri Poincaré discovered that the so-called 3-body problem – the problem of describing the motion of three point masses interacting through Newton's Law of Gravity – exhibits certain degree of "unsolvability". This was in sharp contrast with the view of physicists and mathematicians at the time who in the preceding three centuries since Newton's *Principia* have observed exclusively regular behavior in physical systems. In mathematical terms, "regular behavior" meant that, up to a smooth change of variables, the temporal evolution of a physical system was given by a linear in time flow on "invariant tori" in the phase space of the system.

A resolution of this apparent contradiction was suggested by the Russian mathematician Andrey Kolmogorov in his famous plenary address at the International Congress of Mathematicians held in Amsterdam in 1954 [53], reprinted in [1] (see also [52], reprinted in [54] and [32]). His idea was that while some of the invariant tori are destroyed by perturbations, many invarian tori remain intact, thereby accounting for the "regular" behavior of the system. Kolmogorov's ideas were developed into rigorous mathematical proofs by his student Vladimir Arnold [6] and later by Jürgen Moser [59, 58]. The three initials of their family names were combined to form the acronym KAM under which this circle of results is known today.

These ideas about the behavior of trajectories played a fundamental role in the modern understanding of the deterministic and stochastic behavior of physical systems. On the mathematical side, they have spurred a large amount of rigorous research which is very active to this day. With the advent of computers, researchers started implementing some of the developed rigorous techniques into practical computations. Poincaré's discoveries and KAM-type theorems forever changed the paradigms of classical physics (Aubin and Dahan Dalmedico [8] present an interesting discussion on this topic). The fascinating history of KAM theory is beautifully described (with a minimum of required mathematical background) in the recent book by Dumas [32].

The relative unpopularity of KAM theory among practicing physicists is perhaps due partially to the fact that – while the ideas and implications of KAM theory are not difficult to understand – the rigorous proofs are long and difficult even in their simplest versions, and have very rarely found their way to the pages of physics books for a "general" (but still mathematically oriented) audience. A notable exception is the book by Thirring [72]. More specialized references on the classical KAM theory are Moser [60], Salamon [68] Chierchia [22, 23, 24], Arnold, Kozlov, and Neishtadt [7], Broer and Sevryuk [14], de la Llave [56], Pöschel [64], Benettin *et al* [12].

The standard proofs of KAM-type theorems use the following strategy. The equations that the unknown function should satisfy are complicated and cannot be solved directly. Instead, one performs a sequence of transformations each of which changes the original problem so that after each step one has better control over the solution. The price that we have to pay for this is that we give up some domain, so that after each step we have better control of the problem on a smaller domain. By a judicious choice of the balance between how much the domain is reduced at each step and how much control is gained, one can achieve that the solution found as a limit after infinitely many steps satisfies the original equation over a smaller but non-empty domain. Of course, if the problem has a special structure, in the iterative procedure described above one has to preserve the structure – for example, in KAM theory for Hamiltonian systems one should only use symplectic transformations in order to preserve the form of the Hamilton's equations.

1.1.2 The parameterization method

A fruitful method for some proofs in theory of dynamical systems is the classical graph transform method (see, e.g., [51]). In this method one constructs the desired object (e.g., the stable manifold of a certain map) with some degree of accuracy and then performs a sequence of transformations to make the object closer to the desired one.

Recalling that the goal of KAM theory is to construct a transformation that conjugates the complicated dynamics of the system to some simple dynamics on a torus (e.g., the map on the torus is a translation by a constant amount), we can use the graph transform method with the following purpose. Since we will be working with a symplectic or, more generally, presymplectic flows, let us consider this situation. We are given a symplectic manifold $\mathcal{P}$ and a flow

$$\Phi_t : \mathcal{P} \to \mathcal{P} \ , \qquad t \geq 0 \ ,$$

on the manifold that preserves the symplectic structure. Assume that there exists a submanifold $\mathcal{K}$ of $\mathcal{P}$ that is invariant under time evolution. Let the invariant submanifold $\mathcal{K}$ be topologically an N-dimensional torus, and assume it has certain regularity. We want that the time evolution of the points on the invariant submanifold $\mathcal{K}$,

$$\Phi_t|_{\mathcal{K}} : \mathcal{K} \to \mathcal{K} \ , \qquad t \geq 0 \ ,$$

can then be conjugated to a simple flow on the torus $\mathbb{T}^N$,

$$\phi_t : \mathbb{T}^N \to \mathbb{T}^N : \theta \mapsto \theta + t\omega \ , \qquad t \geq 0 \ ,$$

where ω is a given constant vector whose components are rationally independent, so that the image $\phi_t(\theta)$, $t \in \mathbb{R}$ of any point $\theta \in \mathbb{T}^N$ fills $\mathbb{T}^N$ densely. We can think of the invariant submanifold $\mathcal{K}$ as the image of a map

$$K : \mathbb{T}^N \to \mathcal{P} \ , \qquad \mathcal{K} = K(\mathbb{T}^N) \ ,$$

such that the map K conjugates the flow ϕ_t on $\mathbb{T}^N$ to the flow Φ_t on the invariant submanifold $\mathcal{K} = K(\mathbb{T}^N)$:

$$\Phi_t \circ K = K \circ \phi_t \ .$$

Now the problem becomes to construct the map K. Of course, the map K

is defined up to a constant translation on the torus, but this nonuniqueness is natural and does not create problems.

A natural approach for constructing the conjugating map K is to try to use the geometry and the dynamics of the system. A crucial idea in this direction is that the tangent bundle to the invariant submanifold $\mathcal{K}$ is invariant with respect to the derivative Φ_{t*} of the map Φ_t at the point $k \in \mathcal{K}$, i.e., that

$$(\Phi_{t*})_k\big(T_k\mathcal{K}\big) = T_{\Phi_t(k)}\mathcal{K}\ , \qquad k \in \mathcal{K}\ .$$

This means that, if we choose a basis in the tangent bundle to $\mathcal{P}$ at each point of the invariant submanifold $\mathcal{K}$ in such a way that the first N vectors are tangent vectors to $\mathcal{K}$, then in this basis the matrix representing the derivative $(\Phi_{t*})_k$ will be block upper triangular at each point $k \in \mathcal{K}$. This fact – named "automatic reducibility" – can be used in as an ingredient in proofs as well as to simplify numerical implementations.

1.1.3 The parameterization method in Hamiltonian dynamics

González, Jorba, de la Llave and Villanueva used the geometric ideas described above in their seminal 2005 paper [26] to prove a version of the KAM theorem. They considered a symplectic manifold $\mathcal{P}$ of dimension $2n$ and a map $f : \mathcal{P} \to \mathcal{P}$ preserving the symplectic form and proved the existence of a submanifold $\mathcal{K}$ of dimension n that is invariant with respect to f and such that the map $K : \mathbb{T}^n \to \mathcal{P}$ conjugates the dynamics of f on $\mathcal{K}$ to a translation on $\mathbb{T}^n$. Their proof had an *a posteriori* format, i.e., they assumed the existence of a map $K_0 : \mathbb{T}^n \to \mathcal{P}$ that is only approximate, i.e., that the submanifold $\mathcal{K}_0 = K_0(\mathbb{T}^n)$ is not invariant with respect to the dynamics on $\mathcal{P}$ and that K_0 is only an approximate conjugacy between the translation on $\mathbb{T}^N$ and the dynamics on $\mathcal{K}_0$. Then they showed that under some assumptions, the map K_0 can be used to start an iterative procedure to construct maps K_1, K_2, ..., with $K_j : \mathbb{T}^N \to \mathcal{P}$, for which the errors are smaller, and such that in the limit of $j \to \infty$, the maps K_j tend to the true solution K. The iteration is a version of the Newton method for solving nonlinear equations. This method is convenient to implement in numerical computations. Moreover, *a posteriori* theorems are suitable for validation of numerical results, that is, they can be used to produce computer assisted proof of existence of numerical manifolds.

Another advantage of the method used in [26] is that while the original proofs of KAM theorem relied essentially on using action-angle variables for the system, the parameterization method of [26] does not need action-angle variables at all. This is a big advantage because action-angle variables are often very complicated and/or exhibit singularities.

Methods similar to the ones developed in [26] have been used by González-Enríquez, Haro and de la Llave [41] to study the existence of non-twist tori in degenerate Hamiltonian systems, and by Fontich, de la Llave and Sire [39] and Luque and Villanueva [57] to prove the existence of lower dimensional invariant tori that are partially hyperbolic or elliptic. Since the parameterization method is suitable for efficient numerical implementation, it been used for this purpose by Calleja and de la Llave [15], Huguet, de la Llave and Sire [50], Fox and Meiss [40]. The parameterization method has been the subject of a book published recently by Haro *et al* [48].

1.2 Presymplectic geometry in physics

In modern geometric language, the evolution of an autonomous Hamiltonian system is described as dynamics on the cotangent bundle of the configuration space $\mathcal{N}$ of the system. The cotangent bundle $T^*\mathcal{N}$ has a canonically defined symplectic form Ω, i.e., a non-degenerate closed 2-form. The nondegeneracy of Ω provides an isomorphism between the tangent and the cotangent bundles of the configuration space of the system and is responsible for the existence and uniqueness of the solutions of the Hamilton's equations. In geometric language these are written as

$$\iota_X \Omega = \mathrm{d}H \ .$$

Here X is the vector field governing the dynamics generated by the Hamiltonian H of the system, and ι_X is the contraction with X, i.e., $\iota_X \Omega := \Omega(X, \cdot)$.

If the 2-form Ω is closed and of constant rank but not necessarily non-degenerate, it is called a **presymplectic form**; a manifold endowed with such a form is called a **presymplectic manifold**.

Perhaps the most prominent appearance of presymplectic manifolds is in the transition from Lagrangian to Hamiltonian formalism when certain non-degeneracy conditions are not met. In this section we briefly explain how this happens. We will use temporary notations that are different from the notations used in the rest of the dissertation. Let $\mathcal{N}$ be the configuration space of a mechanical system, with $\dim \mathcal{N} = N$, and let $L : T\mathcal{N} \to \mathbb{R}$ be the Lagrangian of the system. In physics notations, $L = L(q, \dot{q})$, where $q = (q^A)$ are the generalized coordinates and $\dot{q} = (\dot{q}^A)$ are the generalized velocities. The time evolution of the mechanical system is governed by the Euler-Lagrange equations,

$$\frac{\partial L}{\partial q^A} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}^A} = 0 \ , \qquad A = 1, 2, \ldots, N \ . \tag{1.1}$$

To make the transition to Hamiltonian formalism, one defines the generalized momenta,

$$p_A := \frac{\partial L}{\partial \dot{q}^A}(q, \dot{q}) \ , \qquad A = 1, 2, \ldots, N \ . \tag{1.2}$$

The Hamiltonian of the system, $H : T^*\mathcal{N} \to \mathbb{R}$, is defined as the Legendre transformation of L,

$$H(q,p) := \left(\sum_{A=1}^{N} p_A \dot{q}^A - L(q, \dot{q}) \right) \Bigg|_{\dot{q}^A = V^A(q,p)} \ ,$$

where the functions $V^A(q,p)$ are the generalized velocities expressed in terms of q and p from (1.2). According to the Implicit Function Theorem, this is possible exactly when the $N \times N$ matrix $\left[\frac{\partial^2 L}{\partial \dot{q}^A \, \partial \dot{q}^B} \right]$ has maximal rank.

It may well happen, however, that

$$\operatorname{rank} \left[\frac{\partial^2 L}{\partial \dot{q}^A \, \partial \dot{q}^B} \right] = R < N \tag{1.3}$$

– this is the case, e.g., for Lagrangians that depend linearly on some of the velocities $\dot{q}^A$. Let us reorder the coordinates q^A in such a way that the upper left $R \times R$ block of $\left[\frac{\partial^2 L}{\partial \dot{q}^A \, \partial \dot{q}^B} \right]$ is of full rank. Let us also introduce the following notations for the indices: the lowercase roman indices take values from 1 to R, while the lowercase greek indices run from $R+1$ to N:

$$a, b = 1, \ldots, R \ , \qquad \alpha, \beta = R+1, \ldots, N \ .$$

Then the rank condition (1.3) guarantees that we can express the first R velocities $\dot{q}^a$ in terms of coordinates q, the first R momenta p_b, and the remaining $(N-R)$ velocities $\dot{q}^\beta$:

$$\dot{q}^a = V^a(q, p_b, \dot{q}^\beta) \ , \qquad a = 1, \ldots, R \ .$$

Substitute these expressions for $\dot{q}^a$ into (1.2) to obtain

$$\begin{aligned} p_A &= \frac{\partial L}{\partial \dot{q}^A} \,(q, \dot{q}) \,|_{\dot{q}^a = V^a(q, p_b, \dot{q}^\beta)} \\ &= \frac{\partial L}{\partial \dot{q}^A} \left(q, V^a(q, p_b, \dot{q}^\beta), \dot{q}^\alpha\right) \\ &=: \phi_A(q, p_b, \dot{q}^\beta) \ , \qquad A = 1, \ldots, N \ . \end{aligned} \tag{1.4}$$

By the way the functions V^a were obtained, it is clear that $\phi_a(q, p_b, \dot{q}^\beta) = p_a$ for $a = 1, \ldots, N$. In the remaining $(N-R)$ relations (1.4), the functions ϕ_α, $\alpha = R+1, \ldots, N$ cannot depend on $\dot{q}^\beta$ (otherwise we would have been able to express $\dot{q}^\beta$ in terms of the momenta, which would violate the rank condition (1.3)). Therefore the last $(N-R)$ relations from (1.4) are *conditions* on the coordinates and the momenta:

$$p_\alpha = \phi_\alpha(q, p_b) \ , \qquad \alpha = R+1, \ldots, N \ . \tag{1.5}$$

The $(N-R)$ relations (1.5) are called **primary constraints**. They are not dynamical equations, but instead impose $(N-R)$ conditions on the generalized coordinates q and the generalized momenta p, so that they define a subset of the phase space of the system; we assume that this subset is a submanifold

$$\Gamma_1 := \{(q, p) \,|\, p_\alpha = \phi_\alpha(q, p_b) \ , \ \alpha = R+1, \ldots, N\}$$

of dimension

$$\dim \Gamma_1 = 2N - (N - R) = N + R$$

in the $2N$-dimensional phase space of the system.

The above considerations were the starting point of the development of the theory of the so-called constrained systems, initiated by Dirac [28] (see also Dirac's papers [29, 30] and his book [31]) and developed by Bergmann and his collaborators for purposes of quantization of field theories [5, 61, 13]. The book by Sudarshan and Mukunda [69] offers an in-depth exposition of these early works and some later developments.

Clearly, the pull-back Ω_{Γ_1} of the original symplectic form Ω to the submanifold Γ_1 may be degenerate, so that $(\Gamma_1, \Omega_{\Gamma_1})$ is merely a presymplectic manifold. Define the modified Hamiltonian

$$H_c(q, p_b, \dot{q}^\beta) := p_A \dot{q}^A - L(q, \dot{q}) \ ,$$

where in the right-hand side the first R velocities $\dot{q}^a$ are expressed from the first R equations of the system (1.2), and the last $(N - R)$ momenta p_α are given by the constraint equations (1.5). The Hamiltonian vector field X that determines the evolution of the system on the manifold Γ_1 should satisfy

$$(\iota_X \Omega_{\Gamma_1} - \mathrm{d}H_c)|_{\Gamma_1} = 0 \ . \tag{1.6}$$

Since Ω_{Γ_1} is generally presymplectic (i.e., has a nontrivial kernel), the map $X \mapsto \Omega_{\Gamma_1}(X, \cdot)$ is not an isomorphism, so that the equation (1.6) may not have a solution X (think of the completely degenerate case, when the 1-form $\Omega_{\Gamma_1}(X, \cdot)$ is identically zero). To resolve this problem, Gotay, Nester, and Hinds [46] (see also Gotay and Nester [43, 45, 44] and Gotay's Ph.D. thesis [42]) proposed the following iterative procedure. There may exist a set of points in Γ_1 – we assume that this set is a submanifold Γ_2 of Γ_1 – such that the equation (1.6) restricted to Γ_2, i.e.,

$$(\iota_X \Omega_{\Gamma_1} - \mathrm{d}H_c) \circ j_2 = 0 \tag{1.7}$$

has a solution; here $j_2 : \Gamma_2 \hookrightarrow \Gamma_1$ is the natural inclusion.

Although the equation (1.7) has a solution X, it may happen that the vector field X is not tangent to Γ_2. This will imply that Γ_2 is not invariant with with respect to the time evolution, in which case the solution of (1.7) will not have any meaning. Thus, we are forced to look for a subset Γ_3 of Γ_2 (again, assume that Γ_3 is a submanifold of Γ_2) that is invariant with respect to the time evolution. But when we further restrict the dynamics to Γ_3, the new equation,

$$(\iota_X \Omega_{\Gamma_1} - \mathrm{d}H_c) \circ j_3 = 0 \tag{1.8}$$

(where $j_3 : \Gamma_3 \hookrightarrow \Gamma_2$ is the inclusion) may not have a solution because of the non-trivial kernel of the presymplectic form (i.e., for the same reason for which equation (1.6) may not have a solution).

We proceed in this manner to construct a nested sequence of submanifolds

$$\Gamma_1 \supseteq \Gamma_2 \supseteq \cdots \supseteq \Gamma_k \supseteq \Gamma_{k+1} \supseteq \cdots \ .$$

There are three possibilities:

- either there exists some $K \in \mathbb{N}$ for which $\Gamma_K = \emptyset$;
- or the algorithm produces a submanifold $\Gamma_K \neq \emptyset$ with $\dim \Gamma_K = 0$;

- or there exists a K at which the sequence stabilizes, i.e., $\Gamma_K = \Gamma_{K+1}$, and $\dim \Gamma_K \neq 0$.

The first possibility means that the Hamilton equations have no solutions in any sense. In the second case the system is consistent, but the manifold Γ_K consists of isolated points, i.e., it has no dynamics. The most interesting case is the third one, in which we obtain completely consistent equations of motion on the final constraint submanifold Γ_K:

$$(\iota_X \Omega_{\Gamma_1} - \mathrm{d}H_c)\,|_{\Gamma_K} = 0 \ .$$

In general, it may turn out that the pull-back of the symplectic form to the final constraint submanifold Γ_K is degenerate (in particular, nothing prevents Γ_K from being odd-dimensional). Therefore, this construction leads to dynamics on a presymplectic manifold.

Such situations occur in classical electromagnetic theory (see Sec. VI of [46]), in the description of relativistic particles (Hanson, Regge and Teitelboim [47], Sundermeyer [70, Ch. VII]), gauge fields, and generally in systems whose Lagrangians exhibits local symmetries – see, e.g., the books by Sundermeyer [70, 71], Henneaux and Teitelboim [49], Rothe and Rothe [65], the review of Wipf [74], and the philosophical essay of Earman [35]. The above considerations are only the beginning of a long and complicated story which is still unfolding – the book by Henneaux and Teitelboim [49], published more than 20 years ago, has more than 500 pages.

The so-called Dirac brackets developed in connection with constrained dynamics are a standard tool in dealing with constrained systems (see, e.g., the book by Cushman and Bates [25]).

Presymplectic geometry is related to several topics of interest for physicists and mathematicians: equivalence between Lagrangian and Hamiltonian formalisms for constrained systems [18, 11, 10], canonical transformations in presymplectic systems [17, 19], reduction of presymplectic manifolds [27, 37, 36, 63, 3], geometric approach to maximum principles [9], geometric optics [21, 20, 33, 34].

1.3 Goal of the dissertation

The parameterization method for proving KAM theorems [26] has been employed in several contexts, some of which were mentioned at the end of Section 1.1.3. Alishah and de la Llave [4] used this method to prove a KAM theorem for presymplectic systems, when the degeneracy of the presymplectic form causes some problems. They considered a family $\{f_\lambda\}$ of presymplectic maps, i.e., such that each map f_λ from the family preserves the presymplectic form. For such a family they found a value $\bar{\lambda}$ of the parameter λ and an embedding K from a torus to the presymplectic manifold such that

$$f_{\bar{\lambda}} \circ K = K \circ T_\omega$$

where $T_\omega : \theta \mapsto \theta + \omega$ is translation on the torus by a Diophantine vector ω (see Definition 2.9).

The main goal of this dissertation is to prove a KAM theorem for a family $\{V_\lambda\}$ of presymplectic vector fields on an exact presymplectic manifold. In more detail, let us consider an exact presymplectic manifold $(\mathcal{P}, \Omega)$ of dimension $d+2n$, where the kernel of the presymplectic form Ω is d-dimensional; we take $\mathcal{P} \cong \mathbb{T}^d \times T^*\mathbb{T}^n$, where the kernel of Ω coincides with the first d dimensions. Our goal is to find a value $\bar{\lambda}$ of the parameter λ and an embedding $K : \mathbb{T}^{d+n} \to \mathcal{P}$ such that the submanifold $\mathcal{K} := K(\mathbb{T}^{d+n})$ is invariant with respect to the flow Φ_t of the vector field $V_{\bar{\lambda}}$, and K conjugates the flow Φ_t to the linear flow on $\mathbb{T}^{d+n}$,

$$\phi_t : \mathbb{T}^{d+n} \to \mathbb{T}^{d+n} : \theta \mapsto \theta + t\omega \ , \qquad t \geq 0 \ , \tag{1.9}$$

i.e.,

$$\Phi_t \circ K = K \circ \phi_t \ , \qquad t \geq 0 \ . \tag{1.10}$$

We can write (1.10) in infinitesimal form by differentiating (1.10) with respect to t and setting $t = 0$: we obtain that the vector field $V_{\bar{\lambda}}$ at the point $K(\theta)$ should equal the directional derivative $\partial_\omega K$ in the direction of the vector $\omega \equiv \omega_\theta \in T_\theta \mathbb{T}^{d+n}$, i.e.,

$$V_{\bar{\lambda}, K(\theta)} = \partial_\omega K(\theta) \ . \tag{1.11}$$

Here we have used the notation

$$\partial_\omega K(\theta) := (K_*)_\theta \, \omega_\theta \in T_{K(\theta)} \mathcal{P} \ ,$$

where we consider the constant vector ω as a tangent vector to $\mathbb{T}^{d+n}$ at the point $\theta \in \mathbb{T}^{d+n}$. Since we consider $\mathbb{T}^d \times T^*\mathbb{T}^n$ as a model of $\mathcal{P}$, we can think of K as a function with values in $\mathbb{T}^d \times T^*\mathbb{T}^n$, and can assume that there exists a natural basis at $T_p\mathcal{P}$ at each point $p \in \mathcal{P}$, namely, the basis coming from the coordinates in $\mathcal{P} \cong \mathbb{T}^d \times T^*\mathbb{T}^n$. With this understanding, we think of (1.11) as an equality between two column vectors.

The common thread throughout the dissertation will be the geometry of the system and its relation with the dynamics of the presymplectic vector fields. A central role in the construction of K is played by "automatic reducibility" [26] (mentioned in Section 1.1.2), i.e., the fact that one can use the invariance of the tangent bundle to $\mathcal{K}$ with respect to the flow of the presymplectic vector field to construct a special basis in which the equations can be solved simply (see the construction of the corrections ε_0 and Δ_0 below). Another fact that helps us recognize the "big" and "small" parts of certain expressions is that the invariant torus $\mathcal{K}$ is an isotropic submanifold (i.e., that the pull-back of the presymplectic form Ω on $\mathcal{K}$ vanishes identically). We found the following interesting quotations related to this fact. On page 45 of his classic 1973 monograph [60], Moser writes

> Actually, more than asserted in Theorem 2.7 can be proven. It turns out that the differential form $\sum_{\mathrm{k}=1}^{\mathrm{n}} \mathrm{dy_k} \wedge \mathrm{dx_k}$ vanishes identically on the tori (3.11), and one calls manifolds with this property and of maximal dimension Lagrange manifolds.

In this quotation, Theorem 2.7 is (as Moser calls it) the Kolmogorov-Arnold Theorem, and the tori (3.11) are the invariant tori whose existence is proved in the KAM theorem. On page 584 of their monograph [1], Abraham and Marsden write

> Moser [1973a] states that the invariant tori are Lagrangian submanifolds [...]. This fact can probably be exploited, although to our knowledge it has not been.

The fact that the invariant torus $\mathcal{K}$ is isotropic and of a maximum dimension (i.e., Lagrangian in the symplectic case) has been used at a crucial point of the proof in the paper by de la Llave *et al* [26], and is also a vital part of our proof.

We return to the construction of the parameter $\bar{\lambda}$ and the embedding K for which we use the strategy proposed in [26]. Suppose that we are given a value

λ_0 of the parameter λ and a map $K_0 : \mathbb{T}^{d+n} \to \mathcal{P}$ such that the flow $\Phi_{0,t}$ of the vector field V_{λ_0} is approximately conjugate to the linear flow (1.9) on $\mathbb{T}^{d+n}$. Let

$$e_0(\theta) := V_{\lambda_0, K_0(\theta)} - \partial_\omega K_0(\theta) \tag{1.12}$$

be the error. If the pair (λ_0, K_0) were a true solution, then e_0 would be identically zero. We want to construct a more accurate solution (λ_1, K_1), for which the error

$$e_1(\theta) := V_{\lambda_1, K_1(\theta)} - \partial_\omega K_1(\theta) \tag{1.13}$$

would be quadratically small, i.e.,

$$\|e_1\| \leq C \|e_0\|^2 \tag{1.14}$$

(to simplify this explanation, at the moment we ignore the question of choice of norms). To this end, we set

$$\lambda_1 := \lambda_0 + \varepsilon_0 \ , \qquad K_1(\theta) := K_0(\theta) + \Delta_0(\theta) \ . \tag{1.15}$$

and look for the "small" corrections $\varepsilon_0 \in \mathbb{R}^{d+2n}$ and $\Delta_0 : \mathbb{T}^{d+n} \to \mathcal{P}$ to the parameter λ_0 and the embedding K_0, respectively.

The corrections ε_0 and Δ_0 must satisfy a variational equation which is a linear equation with respect to ε_0 and Δ_0. This equation, however, is difficult to solve, so we use the geometry of the system. To understand the geometry, let us go back to the pair $(\bar{\lambda}, K)$. It is clear that the tangent bundle $T\mathcal{K}$ to the invariant submanifold $\mathcal{K} = K(\mathbb{T}^{d+n})$ is an invariant subbundle of the tangent bundle $T\mathcal{P}$ to the manifold $\mathcal{P}$. We can use this fact to construct a special basis in $T_{K(\theta)}\mathcal{P}$ at each point $K(\theta) \in \mathcal{K}$. Namely, we can choose the first $(d+n)$ basis vectors to be tangent to $\mathcal{K}$, and the remaining n basis vectors to be transversal to $\mathcal{K}$. Then the span of the first $(d+n)$ basis vectors will be invariant under the flow Φ_t of the vector field $V_{\bar{\lambda}}$. Then the matrix of Φ_t in this basis will be block upper triangular, with the top left $(d+n) \times (d+n)$ block corresponding to the transformation of the tangent bundle of $\mathcal{K}$.

Using these geometric ideas, we can rewrite the equation for the corrections ε_0 and Δ_0 in the special basis, in which the equation looks simpler up to small corrections. We show that these corrections are small and can be ignored, and then we solve the resulting equation to find ε_0 and Δ_0. Having found ε_0 and Δ_0, we construct λ_1 and Δ_1 according to (1.15), so that the new error, e_1 (1.13), is quadratically small in comparison with the old one, e_0 (1.12), i.e., (1.14) is satisfied. This comes with a price – the functions K_0 and K_1 should be defined on a domain that is the torus $\mathbb{T}^{d+n}$ "thickened" in complex direction, i.e., each angle θ^α is a complex number with $|\mathrm{Im}\ \theta^\alpha| \le \rho$ (see the definition of the thickened domain in (2.12)). The new function, K_1 is defined on a domain $\mathbb{T}^{d+n}_{\rho_1}$ that is smaller than the domain $\mathbb{T}^{d+n}_{\rho_0}$ of K_0, i.e., $\rho_0 > \rho_1$.

We apply the above construction of (λ_1, K_1) iteratively. Namely, we construct a series of pairs (λ_j, K_j) such that the errors e_j (constructed analogously to (1.12) and (1.13)) satisfy

$$\|e_{j+1}\|_{\rho_{j+1}} \le C \, \|e_j\|^2_{\rho_j} \ ,$$

where $\| \ \|_{\rho_j}$ stands for the supremum norm on $\mathbb{T}^{d+n}_{\rho_j}$. Since the functions K_j are defined on domains $\mathbb{T}^{d+n}_{\rho_j}$ that decrease with j, we have to make sure that the decreasing sequence

$$\rho_0 > \rho_1 > \rho_2 > \cdots$$

has a non-zero limit, $\rho_\infty := \lim_{j\to\infty} \rho_j$, so that the function

$$K_\infty := \lim_{j\to\infty} K_j$$

is defined on a nonempty domain $\mathbb{T}^{d+n}_{\rho_\infty}$. To achieve this, one has to carefully choose the balance between how much domain is given up and how much control over the norms is gained; this procedure was introduced in the classic papers by Moser [59, 58].

1.4 Plan of the exposition

In Chapter 2 we define the concepts needed and state the main theorem. In Section 2.1 we introduce exact presymplectic manifolds and presymplectic vector fields and give several conditions for presymplecticity of a vector field. We also discuss some issues related to constructing a symplectic manifold out of a presymplectic one by modding out the kernel of Ω. In Section 2.2 we collect several definitions and state an important result by Rüssmann that is used later. In Section 2.3 we set up the problem, introduce notations for the coordinates, and give a complete statement of our main result, Theorem 2.12.

In Chapter 3 we discuss the geometry of the problem, assuming that we know a true solution $K : \mathbb{T}^{d+n} \to \mathcal{P}$ satisfying (1.10). After giving some definitions in Section 3.1, we prove in Section 3.2 that an invariant torus is isotropic. Section 3.3 is devoted to the construction of an adapted basis in the linear spaces $T_{K(\theta)}\mathcal{P}$ for each $K(\theta) \in \mathcal{K}$ (i.e., a basis of $(T\mathcal{P})|_{\mathcal{K}}$) with the special property that the first $(d+n)$ vectors in it span $T_{K(\theta)}\mathcal{K}$ and writing the presymplecticity condition on the vector field V_λ in this adapted basis. In Section 3.4 we construct a matrix

M_θ of change of basis from an arbitrary basis in $T_{K(\theta)}\mathcal{P}$ to the adapted basis. We also derive a representation of $(DV_{K(\theta)} - \partial_\omega)M_\theta$ which will be useful in solving the linearized equation in Section 4.3 and find expressions for M_θ^{-1}.

Chapter 4 is devoted to computing the approximate solutions (λ_j, K_j) from the initial approximation (λ_0, K_0). For an approximate solution K_0, the torus $\mathcal{K}_0 = K_0(\mathbb{T}^{d+n})$ is not an isotropic submanifold of $\mathcal{P}$, but the norm of the pull-back of the presymplectic form Ω to $\mathcal{K}_0$ is small and can be bounded above by the norm of the error; we derive these bounds in Section 4.1. In Section 4.2 we derive the variational equation whose solutions are the corrections ε_0 and Δ_0 to λ_0 and K_0 (recall (1.15)). We use the adapted basis constructed in Section 3.3 to identify the "big" and the "small" parts of the coefficients in the variational equation. We ignore the "small" parts to write a simplified version of the variational equation in which the terms that were ignored are the same order as the terms that were neglected in the derivation of the variational equation (so that ignoring the "small" terms does not contribute to the leading order of the error). We solve the resulting equation, thus finding the corrections ε_0 and Δ_0 to the parameter λ_0 and the embedding K_0.

Having shown how to correct (λ_0, K_0) to construct a better approximation (λ_1, K_1), we apply this procedure iteratively to construct a sequence of approximations

$$(\lambda_0, K_0) \mapsto (\lambda_1, K_1) \mapsto (\lambda_2, K_2) \mapsto (\lambda_3, K_3) \mapsto \cdots$$

whose limit $(\lambda_\infty, K_\infty)$ is the desired solution $(\bar{\lambda}, K)$ satisfying (1.10). This iterative construction should be done carefully so that the domain of the limiting embedding K_∞ is non-empty. This is performed in Chapter 5.

Chapter 2

Preliminaries and General Setup

In this chapter, we will introduce the definitions and notations for our setup. In particular, we will define presymplectic manifolds, presymplectic vector fields, Diophantine vectors, and describe the classes of functions and norms that we will be using. We will also see that there exists a canonical symplectic manifold obtained from the presymplectic manifold given by modding out by the kernel of the presymplectic form. We introduce coordinates adapted to the geometry of the problem, and give a precise statement of the main theorem of this dissertation.

2.1 Exact presymplectic manifolds

The systems that we will be considering are presymplectic vector fields on exact presymplectic manifolds.

Definition 2.1. *A* ***presymplectic manifold*** *is a pair* $(\mathcal{P}, \Omega)$*, where* $\mathcal{P}$ *is a manifold of any (finite) dimension and* $\Omega \in \Omega^2(\mathcal{P})$ *is a closed 2-form with constant rank. If* Ω *is exact, i.e., if* $\Omega = \mathrm{d}\tau$ *for some* $\tau \in \Omega^1(\mathcal{P})$*, then we say that* $(\mathcal{P}, \Omega)$ *is an* ***exact presymplectic manifold****.*

Throughout this dissertation, we will always assume that

$$\dim \mathcal{P} = d + 2n \ , \qquad \text{rank } \Omega = 2n \ . \tag{2.1}$$

Most of the time we will consider the specific exact presymplectic manifold

$$\mathcal{P} := \mathbb{T}^d \times T^*\mathbb{T}^n \cong \mathbb{T}^d \times \mathbb{T}^n \times \mathbb{R}^n \tag{2.2}$$

with an exact presymplectic form Ω of rank $2n$ whose kernel is coincides with the d-dimentional torus, $\mathbb{T}^d$. We will assume that the manifold $\mathbb{T}^d \times T^*\mathbb{T}^n$ is endowed with an Euclidean structure, so that we can identify two-forms with linear operators and abstract tangent vectors with column vectors. This will be useful for doing analysis in later chapters. Choosing $\mathbb{T}^d \times T^*\mathbb{T}^n$ as a model for a general presymplectic manifold is a natural choice employed by many researchers in the field.

Despite the specific choice (2.2) of the structure of the exact presymplectic manifold $\mathcal{P}$, we will use general differential-geometric ideas as an inspiration. In Section 2.1.2 we will discuss some general differential-geometric aspects of the problem at hand without using the specific structure of $\mathcal{P}$ given by (2.2).

One important way in which a presymplectic manifold differs from a symplectic manifold is that a symplectic manifold must be even dimensional, whereas a presymplectic manifold could be of even or odd dimension. If the rank of Ω is equal to the dimension of $\mathcal{P}$ (and $\dim \mathcal{P}$ is even), then the manifold is actually a symplectic manifold.

2.1.1 Presymplectic vector fields

Given a presymplectic manifold $\mathcal{P}$, we will consider a certain class of vector fields on $\mathcal{P}$. In the definition below, $\mathfrak{X}(\mathcal{P})$ stands for the vector fields on $\mathcal{P}$, $\mathcal{L}$ is the Lie Derivative, and ι is the interior product, i.e., the contraction of a form with a vector field.

Definition 2.2. *Let $V \in \mathfrak{X}(\mathcal{P})$ be a vector field on $\mathcal{P}$ and*

$$\Phi_t \colon \mathcal{P} \to \mathcal{P}$$

be the time-t flow of V. The vector field V is said to be ***presymplectic*** *if its flow Φ_t preserves the presymplectic structure on $\mathcal{P}$, i.e. if*

$$\Phi_t^* \, \Omega = \Omega \qquad \forall \, t \in \mathbb{R} \; .$$

The following proposition gives several equivalent conditions for a vector field to be presymplectic.

Proposition 2.3. *Let $(\mathcal{P}, \Omega)$ be a presymplectic manifold, and $V \in \mathfrak{X}(\mathcal{P})$. Then the following conditions are equivalent:*

(a) V is a presymplectic vector field;

(b) the Lie derivative of the presymplectic form along V vanishes:

$$\mathcal{L}_V \Omega = 0 \; ; \tag{2.3}$$

(c) the 1-form $\iota_V \Omega$ is closed.

Proof. The fact that (a) implies (b) follows directly from the definition of a Lie derivative:

$$\mathcal{L}_V \Omega = \lim_{t\to 0} \frac{\Phi_t^* \Omega - \Omega}{t} \ .$$

To see that (b) implies (a), notice that

$$\Phi_t^* \left(\mathcal{L}_V \Omega \right) = \Phi_t^* \left(\lim_{s\to 0} \frac{\Phi_s^* \Omega - \Omega}{s} \right) = \lim_{s\to 0} \frac{\Phi_{t+s}^* \Omega - \Phi_t^* \Omega}{s} = \frac{\mathrm{d}}{\mathrm{d}s} (\Phi_s^* \Omega) \Big|_{s=t} \ .$$

This observation implies that, if $\mathcal{L}_V \Omega = 0$, then $\Phi_t^* \Omega$ is constant for all t, hence

$$\Phi_t^* \Omega = \Phi_0^* \Omega = \Omega \ .$$

To show that (b) and (c) are equivalent, we use Cartan's magic formula and the closedness of Ω:

$$\mathcal{L}_V \Omega = \iota_V \mathrm{d}\Omega + \mathrm{d}(\iota_V \Omega) = \mathrm{d}(\iota_V \Omega) \ .$$

$\square$

We will be interested in families of vector fields, $\{V_\lambda\}$, where λ is a parameter in $\mathbb{R}^m$ for some m which is usually equal to $\dim \mathcal{P}$. Considering families of vector fields instead of just a single vector field means that we will be looking for a particular value of the parameter λ, say $\bar{\lambda}$, for which the vector field $V_{\bar{\lambda}}$ from the family $\{V_\lambda\}$ has an invariant torus. In other words, in the course of our computations we will need to adjust the parameter λ to a value $\bar{\lambda}$ for which our problem has a solution. This is a commonly used technical procedure (see, e.g., [26, 38, 4, 16]) that will allow us to have a more relaxed set of non-degeneracy conditions. We will often suppress the subscript λ to help make formulae more visually pleasing, especially when the value of the parameter λ does not change.

2.1.2 Foliation induced by $\ker\Omega$

In this section we consider more general questions related to the geometry of presymplectic manifolds. Since in our proof we will work only in the case when $\mathcal{P}$ has the concrete form given in (2.2), this subsection is mostly of theoretical interest.

For any $p \in \mathcal{P}$, the presymplectic form

$$\Omega_p : T_p\mathcal{P} \times T_p\mathcal{P} \to \mathbb{R}$$

is an antisymmetric bilinear mapping, and its kernel is defined as

$$\begin{aligned}\ker\Omega_p &:= \{W_p \in T_p\mathcal{P} \mid \iota_{W_p}\Omega_p = 0\} \\ &= \{W_p \in T_p\mathcal{P} \mid \Omega_p(W_p, U_p) = 0 \ \ \forall\, U_p \in T_p\mathcal{P}\} \subseteq T_p\mathcal{P}\ .\end{aligned}$$

Since Ω is of constant rank $2n$ and $\dim\mathcal{P} = d + 2n$ (recall (2.1)), the collection of subspaces $\ker\Omega_p$ for all $p \in \mathcal{P}$ forms a differentiable distribution of constant rank d, i.e.,

$$\dim\ker\Omega_p = d\ , \qquad p \in \mathcal{P}\ ,$$

which we denote by $\ker\Omega$. Let $\mathfrak{X}^{\ker\Omega}(\mathcal{P})$ stand for the set of all smooth vector fields on $\mathcal{P}$ whose value at each point $p \in \mathcal{P}$ lies in $\ker\Omega_p$:

$$\begin{aligned}\mathfrak{X}^{\ker\Omega}(\mathcal{P}) &:= \{W \in \mathfrak{X}(\mathcal{P}) \mid \iota_W\Omega = 0\} \\ &= \{W \in \mathfrak{X}(\mathcal{P}) \mid W_p \in \ker\Omega_p\ \forall\, p \in \mathcal{P}\}\ .\end{aligned}$$

One of the important immediate consequences from the fact that Ω is presymplectic is that its kernel is a completely integrable distribution. To prove this result, we will need the classical Frobenius Theorem, proved in the books of Warner [73, Theorem 1.60], Abraham and Marsden [2, Section 4.4], Rudolph and Schmidt [66, Section 3.5], and Libermann and Marle [55, Appendix 3], among many others; Libermann and Marle give a very detailed discussion.

Theorem 2.4 (Frobenius). *On a manifold $\mathcal{M}$, let $\mathscr{F}$ be a differentiable distribution of constant rank. Then $\mathscr{F}$ is completely integrable if and only if for every pair (U, V) of differentiable sections of $\mathscr{F}$, defined on the same open subset of $\mathcal{M}$, $[U, V]$ is a differentiable section of $\mathscr{F}$.*

Lemma 2.5. *If Ω is a presymplectic form, the distribution $\ker \Omega$ is completely integrable.*

Proof. Let $U, V \in \mathfrak{X}^{\ker \Omega}(\mathcal{P})$ and let $W \in \mathfrak{X}(\mathcal{P})$ be an arbitrary vector field on $\mathcal{P}$. From the closedness of Ω and the explicit formula for exterior derivatives we obtain

$$\begin{aligned} 0 &= (\mathrm{d}\Omega)(U, V, W) \\ &= U(\Omega(V, W)) - V(\Omega(U, W)) + W(\Omega(U, V)) \\ &\quad - \Omega([U, V], W) + \Omega([U, W], V) - \Omega([V, W], U) \\ &= -\Omega([U, V], W)\ , \end{aligned}$$

hence $[U, V] \in \mathfrak{X}^{\ker \Omega}(\mathcal{P})$. The integrability of $\ker \Omega$ follows directly from this observation and the Frobenius Theorem. $\square$

Lemma 2.5 implies that the manifold $\mathcal{P}$ has a foliation with d-dimensional leaves such that the tangent space to the leaf through a point $p \in \mathcal{P}$ is the subspace $\ker \Omega_p$ of $T_p\mathcal{P}$. Clearly, even if a distribution is completely integrable, the leaves of the resulting foliation might not form a manifold. To formulate a condition that guarantees that the leaves of the foliation form a manifold we need the following definition (reproduced from [55, Sec. 4.3.3 of Appendix 3]).

Definition 2.6. *A foliation on a differentiable manifold $\mathcal{P}$ is said to be **simple** if there exists a surjective submersion $\pi^{\mathcal{Q}}$ of $\mathcal{P}$ onto another differentiable manifold $\mathcal{Q}$ such that, for every point p in $\mathcal{P}$, the leaf that passes through p is the closed submanifold $\left(\pi^{\mathcal{Q}}\right)^{-1}(\pi^{\mathcal{Q}}(p))$. The manifold $\mathcal{Q}$ may then be identified with the set of leaves of the foliation, and the distribution tangent to the leaves is* $\ker\left(\pi^{\mathcal{Q}}_*\right)$, *where $\pi^{\mathcal{Q}}_*$ stands for the derivative of the map $\pi^{\mathcal{Q}}$.*

The proposition below (adapted from [55, Section III.7]) states that the manifold $\mathcal{Q}$ carries a natural symplectic structure.

Proposition 2.7. *Let $(\mathcal{P}, \Omega)$ be a presymplectic manifold with $\operatorname{rank} \Omega = 2n$. Assume that the foliation defined by the completely integrable distribution* $\ker \Omega$ *of $T\mathcal{P}$ is simple.*

Let $\mathcal{Q}$ be the manifold of the leaves of this foliation, and

$$\pi^{\mathcal{Q}} \colon \mathcal{P} \to \mathcal{Q} \tag{2.4}$$

be the canonical projection. Then there exists a unique symplectic form $\widetilde{\Omega} \in \Omega^2(\mathcal{Q})$ on the manifold $\mathcal{Q}$ such that

$$(\pi^{\mathcal{Q}})^*\widetilde{\Omega} = \Omega \ . \tag{2.5}$$

Definition 2.8. *The symplectic manifold $(\mathcal{Q}, \widetilde{\Omega})$ constructed in Proposition 2.7* *is called the* ***reduced symplectic manifold*** *associated with the presymplectic manifold* $(\mathcal{P}, \Omega)$.

For the remainder of the dissertation, we will always assume that the foliation induced by $\ker\Omega$ is simple and, therefore, the collection of leaves is a manifold $\mathcal{Q}$. Clearly, this assumption is satisfied in the particular case (2.2) of main interest for us. It would be interesting to investigate the case when this condition is not met.

2.1.3 Matrix representation of Ω and $\widetilde{\Omega}$

Since the kernel of the symplectic form Ω in the presymplectic manifold $\mathcal{P}$ (2.2) is assumed to coincide with $\mathbb{T}^d$, the collection of leaves,

$$\mathcal{Q} = \mathcal{P}/\ker\Omega = T^*\mathbb{T}^n \ , \tag{2.6}$$

is a symplectic manifold with symplectic form $\widetilde{\Omega}$ given by (2.5). Since we assume the existence of Euclidean structure on $\mathcal{P}$ (recall (2.2)), we can identify a 2-form with a linear operator. Let

$$J_p : T_p\mathcal{P} \to T_p\mathcal{P} \ , \qquad p \in \mathcal{P}$$

be the linear operator corresponding to the presymplectic form Ω on $\mathcal{P}$ at $p \in \mathcal{P}$, which is defined by

$$\langle U_p, J_p W_p\rangle_{\mathbb{R}^{d+2n}} = \Omega_p(U_p, W_p) \ , \qquad U_p, W_p \in T_p\mathcal{P} \cong \mathbb{R}^{d+2n} \ , \quad p \in \mathcal{P} \ , \tag{2.7}$$

where $\langle \cdot, \cdot \rangle_{\mathbb{R}^{d+2n}}$ is the Euclidean inner product on $\mathbb{R}^{d+2n}$. Similarly, let

$$\tilde{J}_q : T_q\mathcal{Q} \to T_q\mathcal{Q} \ , \qquad q \in \mathcal{Q}$$

be the linear operator corresponding to the symplectic form $\widetilde{\Omega}$ on $T_q\mathcal{Q}$ at $q \in \mathcal{Q}$, defined by

$$\left\langle \widetilde{U}_q, \widetilde{J}_q \widetilde{W}_q \right\rangle_{\mathbb{R}^{2n}} = \widetilde{\Omega}_q\Big(\widetilde{U}_q, \widetilde{W}_q\Big) \ , \qquad \widetilde{\xi}, \widetilde{\eta} \in T_q\mathcal{Q} \cong \mathbb{R}^{2n} \ , \quad q \in \mathcal{Q} \ , \tag{2.8}$$

where $\langle \cdot, \cdot \rangle_{\mathbb{R}^{2n}}$ is the Euclidean inner product on $\mathbb{R}^{2n}$.

Since $\widetilde{\Omega}$ is a symplectic form, it is clear that $\widetilde{J}_q$ is a linear isomorphism for any $q \in \mathcal{Q}$. On the other hand, J_p has a d-dimensional kernel. If we choose a basis for $T_p\mathcal{P} \cong \mathbb{R}^d \times \mathbb{R}^{2n}$ such that the first d vectors form a basis of $\mathbb{R}^d$, and the other $2n$ vectors form a basis of $\mathbb{R}^{2n}$, then we can write J_p in a matrix form as

$$J_p = \begin{bmatrix} 0 & 0 \\ 0 & \widetilde{J}_{\pi^{\mathcal{Q}}(p)} \end{bmatrix} \ . \tag{2.9}$$

We will not make a notational distinction between an operator and its matrix.

Clearly, the antisymmetry of Ω and $\widetilde{\Omega}$ imply the antisymmetry of J_p and $\widetilde{J}_q$:

$$J_p^\top = -J_p \ , \qquad \widetilde{J}_q^\top = -\widetilde{J}_q \ .$$

Although J_p is not invertible, we will use the notation J_p^{-1} for the Moore-

Penrose pseudoinverse for J_p:

$$J_p^{-1} := \begin{bmatrix} 0 & 0 \\ 0 & \widetilde{J}_{\pi^{\mathcal{Q}}(p)}^{-1} \end{bmatrix} \; ; \tag{2.10}$$

with this definition we have

$$J_p \, J_p^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{I}_{2n} \end{bmatrix} \; .$$

2.2 Miscellaneous definitions and results

In this section we collect several definitions and results that will be needed later.

2.2.1 Diophantine vectors

Diophantine numbers hold a special role in mathematics. These numbers are (necessarily) irrational, but in some sense they are "more irrational" than some other irrational numbers. The properties of these numbers help to overcome the problem of "small divisors" and thus to solve a certain linear differential equation on the torus (as in Proposition 2.11 below).

Definition 2.9. *For $\gamma > 0$ and $\sigma \geq d+n-1$, the set of all $\omega \in \mathbb{R}^{d+n}$ satisfying the condition*

$$|\omega \cdot k| \geq \frac{\gamma}{|k|^{\sigma}} \qquad \forall \; k \in \mathbb{Z}^{d+n} \setminus \{0\} \tag{2.11}$$

will be called the set of ***Diophantine vectors*** *and will be denoted by $\mathcal{D}(\gamma, \sigma)$.*

Diophantine vectors are abundant, being of full measure (in the Lebesgue sense). For an in-depth look at the properties of Diophantine numbers see, e.g., Niven's book [62].

2.2.2 Function spaces

For any $\rho > 0$, we define the torus "thickened" into the complex direction,

$$\mathbb{T}_\rho^{d+n} := \{\theta \in \mathbb{C}^{d+n}/\mathbb{Z}^{d+n} \mid |\mathrm{Im}\ \theta^\alpha| \leq \rho,\ \alpha = 1, 2, \ldots, d+n\}\ . \tag{2.12}$$

Let $|\ |$ stand for the supremum norm on $\mathbb{R}^m$ or $\mathbb{C}^m$ (for any m). Given $\rho > 0$, we define the set of functions $\mathcal{W}_\rho$ as follows:

$$\begin{aligned}\mathcal{W}_\rho := \Big\{K : \mathbb{T}_\rho^{d+n} \to \mathcal{P} \mid\ &\text{(a) } K \text{ is real analytic on the interior of } \mathbb{T}_\rho^{d+n}\ ,\\ &\text{(b) } K \text{ is continuous on the boundary of } \mathbb{T}_\rho^{d+n}\ ,\ \text{and}\\ &\text{(c) } K \text{ is periodic of period 1 in all of its arguments}\Big\}\ .\end{aligned} \tag{2.13}$$

Define a norm on $\mathcal{W}_\rho$ by

$$\|K\|_\rho = \sup_{\theta \in \mathbb{T}_\rho^{d+n}} |K(\theta)|\ .$$

With the above definitions, $(\mathcal{W}_\rho\, , \|\ \|_\rho)$ is a Banach space.

We will use also the following norms: for analytic functions g with bounded derivatives in a complex domain $\mathcal{B}$ and for $\ell \in \mathbb{N}$,

$$|g|_{C^\ell, \mathcal{B}} := \sup_{0 \leq |\mathsf{k}| \leq \ell}\ \sup_{z \in \mathcal{B}} \left|D^{\mathsf{k}} g(z)\right|\ . \tag{2.14}$$

where k is a multiindex.

The following bound is an easy application of the Cauchy integral formula.

Proposition 2.10. *For $K \in \mathcal{W}_\rho$ and $0 < \delta < \rho$, the following inequality holds:*

$$\|DK\|_{\rho-\delta} \leq C\delta^{-1} \|K\|_\rho \ . \tag{2.15}$$

From the estimate (2.15) we can see that in order to get an estimate on the derivative DK, we must shrink the width of the thickened torus from ρ to $\rho - \delta$. If we select a very small $\delta > 0$, then δ^{-1} is large and we have lost some control over the tightness of the bound. On the other hand, if δ is large, then we have a tight bound in a small domain. Choosing an appropriate δ will play an important role in the convergence of the iterative method.

2.2.3 Rüssmann's result

The main reason for us to consider the set $\mathcal{D}(\gamma, \sigma) \subset \mathbb{R}^{d+n}$ is because of the following proposition by Rüssmann [67].

Proposition 2.11. *Let $\omega = \begin{bmatrix}\omega^1 \ \omega^2 \ \cdots \ \omega^{d+n}\end{bmatrix}^\top \in \mathcal{D}(\gamma, \sigma)$ and let the function $h : \mathbb{T}^{d+n} \to \mathcal{P}$ be analytic on $\mathbb{T}^{d+n}_\rho$ and have zero average. Then for any $0 < \delta < \rho$, the differential equation*

$$\partial_\omega v = h \ ,$$

where

$$\partial_\omega := \omega^1 \frac{\partial}{\partial \theta^1} + \cdots + \omega^{d+n} \frac{\partial}{\partial \theta^{d+n}}$$

is the directional derivative in the direction of ω, has a unique average zero solution $v : \mathbb{T}^{d+n} \to \mathcal{P}$ which is analytic in $\mathbb{T}^{d+n}_{\rho-\delta}$.

Moreover, the solution v satisfies the estimate

$$\|v\|_{\rho-\delta} < C\gamma^{-1}\delta^{-\sigma} \|h\|_\rho \ , \tag{2.16}$$

where C is a constant depending only on d, n, and σ.

2.3 Setting up the problem

In this section we describe briefly the general setup of the problem, introduce some notations, and state our main result (Theorem 2.12).

2.3.1 General setup

Let $V_\lambda \in \mathfrak{X}(\mathcal{P})$ be a $(d+2n)$-parameter family of presymplectic vector fields on the exact presymplectic manifold $\mathcal{P}$. The goal is to construct a torus in $\mathcal{P}$ that is invariant with respect to the flow of the vector field V_λ for some value $\bar{\lambda}$ of the parameter λ and, moreover, such that the flow of $V_{\bar{\lambda}}$ on this invariant torus be conjugate to a translation on $\mathbb{T}^{d+n}$ by a Diophantine vector ω. More specifically, let $\mathbb{T}^{d+n} := \mathbb{R}^{d+n}/\mathbb{Z}^{d+n}$ be the $(d+n)$-dimensional torus, and

$$K : \mathbb{T}^{d+n} \to \mathcal{P} \tag{2.17}$$

be a smooth embedding. We want that the torus

$$\mathcal{K} := K(\mathbb{T}^{d+n}) \subseteq \mathcal{P} \tag{2.18}$$

(of dimension $d+n$) be invariant with respect to the flow Φ_t of the presymplectic vector field $V_{\bar{\lambda}}$, and, moreover, that the flow Φ_t of the vector field $V_{\bar{\lambda}}$ be conjugate to the linear flow ϕ_t along the constant Diophantine vector $\omega \in \mathbb{R}^{d+n}$:

$$K(\phi_t(\theta)) = \Phi_t(K(\theta)) \qquad \forall\, t \in \mathbb{R}\ , \quad \forall\, \theta \in \mathbb{T}^{d+n}\ , \tag{2.19}$$

where

$$\phi_t : \mathbb{T}^{d+n} \to \mathbb{T}^{d+n} : \theta \mapsto \theta + t\omega \ . \tag{2.20}$$

When we write an argument of K, we always think of it as an element of $\mathbb{T}^{d+n}$, without writing this explicitly (i.e., when writing $\theta + t\omega$, we assume that we have taken only the fractional parts of each component of $\theta + t\omega$).

We can express the relation (2.19) by saying that the diagram

$$\begin{tikzcd}
\mathbb{T}^{d+n} \arrow[r, "K"] \arrow[d, "\phi_t"'] & \mathcal{P} \arrow[d, "\Phi_t"] \\
\mathbb{T}^{d+n} \arrow[r, "K"] & \mathcal{P}
\end{tikzcd}$$

is commutative.

Taking a derivative with respect to t of both sides of (2.19) and setting $t = 0$, we obtain that (2.19) implies

$$K_{*\,\theta}\, \omega_\theta = V_{\bar{\lambda}, K(\theta)} \qquad \forall\, \theta \in \mathbb{T}^{d+n} \ . \tag{2.21}$$

Here $V_{\bar{\lambda},K(\theta)} \in T_{K(\theta)}\mathcal{K} \subseteq T_{K(\theta)}\mathcal{P}$ is the value of the vector field $V_{\bar{\lambda}}$ at the point $K(\theta) \in \mathcal{K} \subseteq \mathcal{P}$, $\omega_\theta \in T_\theta \mathbb{T}^{d+n}$ is the Diophantine vector ω considered as an element of the tangent space $T_\theta\mathbb{T}^{d+n}$ (which is naturally isomorphic to $\mathbb{R}^{d+n}$), and

$$K_{*\,\theta} : T_\theta \mathbb{T}^{d+n} \to T_{K(\theta)}\mathcal{K} \subseteq T_{K(\theta)}\mathcal{P}$$

is the derivative of the map K (2.17) at the point $\theta \in \mathbb{T}^{d+n}$. Hereafter, we usually write the arguments as subscripts.

2.3.2 Notations for the coordinates

Instead of the differential-geometric notations used in (2.21), we will normally use matrix notations, and will write the arguments as subscripts. In these notations the equality (2.21) reads

$$DK_\theta\,\omega = V_{\bar\lambda,K(\theta)} \tag{2.22}$$

(where ω is considered as a constant column vector) or, using ∂_ω for directional derivative in the direction of ω,

$$\partial_\omega K_\theta = V_{\bar\lambda,K(\theta)}\ .$$

In (2.22), DK_θ stands for the matrix

$$DK_\theta = \left[\left(DK_\theta\right)^A{}_\alpha\right] = \left[\frac{\partial K^A}{\partial\theta^\alpha}(\theta)\right] \in \mathrm{M}_{d+2n,d+n}(\mathbb{R})\ .$$

Hereafter we use the following notations for the indices: capital roman letters stand for the coordinates in $\mathcal{P}$, while lowercase letters from the beginning of the greek alphabet index the coordinates in $\mathbb{T}^{d+n}$:

$$A, B, \ldots = 1, 2, \ldots, d+2n\ , \qquad \alpha, \beta, \ldots = 1, 2, \ldots, d+n\ . \tag{2.23}$$

Writing $\mathcal{P}$ as $\mathbb{T}^d \times T^*\mathbb{T}^n$ as in (2.2), we divide the coordinates $x = \left(x^A\right)$ in two groups: $\underline{x} = (\underline{x}^\mu)$ parameterize $\mathbb{T}^d$ (the first d coordinates in $\mathcal{P}$), while $\widetilde{x} = (\widetilde{x}^i)$ parameterize $T^*\mathbb{T}^n$ (the last $2n$ coordinates in $\mathcal{P}$):

$$x = \left(x^A\right) = (\underline{x}, \widetilde{x}) = \left(\underline{x}^\mu, \widetilde{x}^i\right)\ , \qquad \mu = 1, 2, \ldots, d\ , \quad i = 1, 2, \ldots, 2n\ . \tag{2.24}$$

These notations are collected in Table 2.1 below.

Coordinates	Range of indices	Remark
$x = \left(x^A\right)$	$A, B = 1, 2, \ldots, d + 2n$	Coordinates in $\mathcal{P} \cong \mathbb{T}^d \times T^*\mathbb{T}^n$
$\underline{x} = (\underline{x}^\mu)$	$\mu, \nu = 1, 2, \ldots, d$	Coordinates in $\mathbb{T}^d$ (the first d in $\mathcal{P}$)
$\widetilde{x} = (\widetilde{x}^i)$	$i, j = 1, 2, \ldots, 2n$	Coordinates in $T^*\mathbb{T}^n$ (the last $2n$ in $\mathcal{P}$)
$\theta = (\theta^\alpha)$	$\alpha, \beta = 1, 2, \ldots, d + n$	Coordinates in $\mathbb{T}^{d+n}$

Table 2.1: Notations for indices and coordinates.

2.3.3 The presymplecticity condition in matrix notations

In this section we will use the notations introduced above to write down the condition for a vector field to be presymplectic. The parameter λ plays no role here, so we omit it.

Using the explicit expression for the Lie derivative of a 2-form,

$$(\mathcal{L}_V\Omega)(U,W) = \mathcal{L}_V\big(\Omega(U,W)\big) - \Omega\left(\mathcal{L}_V U, W\right) - \Omega\left(U, \mathcal{L}_V W\right)$$

(where $U, V, W \in \mathfrak{X}(\mathcal{P})$), and the fact that the Lie derivative of a vector field is the commutator

$$(\mathcal{L}_V U)^A = [V,U]^A = \sum_{A=1}^{d+2n} \left(\frac{\partial U^A}{\partial x^B}\, V^B - \frac{\partial V^A}{\partial x^B}\, U^B \right) \ ,$$

we obtain

$$(\mathcal{L}_V\Omega)(U,W) = \sum_{A,B,C=1}^{d+2n} U^A \left(\frac{\partial \Omega_{AB}}{\partial x^C}\, V^C + \frac{\partial V^C}{\partial x^A}\, \Omega_{CB} + \Omega_{AC}\, \frac{\partial V^C}{\partial x^B} \right) W^B \ .$$

Using the operator $J = \big(J^A{}_B\big)$ introduced in (2.7), we have

$$\Omega(U,W) = \langle U, JW\rangle = U^\top JW = \sum_{A,B=1}^{d+2n} U_A J^A{}_B W^B \ ,$$

so we can rewrite the above identity as

$$\begin{aligned}(\mathcal{L}_V\Omega)(U,W) &= \sum_{A,B,C=1}^{d+2n} U_A \left(\frac{\partial J^A{}_B}{\partial x^C}\, V^C + \frac{\partial V^C}{\partial x^A}\, J^C{}_B + J^A{}_C\, \frac{\partial V^C}{\partial x^B}\right) W^B \\ &= \sum_{A,B,C=1}^{d+2n} U_A \left(\frac{\partial J^A{}_B}{\partial x^C}\, V^C + \big((DV)^\top\big)^A{}_C\, J^C{}_B + J^A{}_C\, (DV)^C{}_B\right) W^B \\ &= \sum_{A,B,C=1}^{d+2n} U_A \Big((DJ)\, V + (DV)^\top\, J + J\, DV\Big)^A{}_B\, W^B \ .\end{aligned}$$

Here we lowered an index of a vector to signify transposition (which is the same as contracting the vector with the Euclidean metric tensor), and used the notations

$$(DV)^C{}_B = \frac{\partial V^C}{\partial x^B} \ , \qquad \big((DV)^\top\big)^A{}_C = \frac{\partial V^C}{\partial x^A} \ ,$$

and

$$\Big((DJ)\, V\Big)^A{}_B := \sum_{C=1}^{d+2n} \frac{\partial J^A{}_B}{\partial x^C}\, V^C \ . \tag{2.25}$$

Therefore in matrix notations the condition (2.3) for the vector field V to be presymplectic becomes

$$(DJ)\, V + (DV)^\top\, J + J\, DV = 0 \ . \tag{2.26}$$

2.3.4 Statement of the main theorem

Here we finally give a complete statement of the main theorem in this dissertation.

Theorem 2.12. *Assume that:*

1) $\omega \in \mathcal{D}(\gamma, \sigma)$ *is a Diophantine vector;*

2) $\mathcal{P} = \mathbb{T}^d \times T^*\mathbb{T}^n$*;*

3) Ω *is an exact presymplectic form on* $\mathcal{P}$ *of rank* $2n$ *such that the kernel of* Ω *coincides with the first* d *directions;*

4) $\{V_\lambda\}$ *is a* $(d+2n)$*-parameter family of presymplectic vector fields on* $\mathcal{P}$*;*

5) $K_0 : \mathbb{T}^{d+n} \to \mathcal{P}$ *is an embedding belonging to the class* $\mathcal{W}_{\rho_0}$ (2.13)*;*

6) *the value* λ_0 *of the parameter* λ *is such that the pair* (λ_0, K_0) *is non-degenerate in the sense of Definition 4.6;*

7) *each vector field from the family* $\{V_\lambda\}$ *can be holomorphically extended to some complex neighborhood* $\mathcal{B}_r$ *of* $K_0(\mathbb{T}^{d+n}_\rho)$*, where*

$$\mathcal{B}_r := \left\{ z \in \mathbb{C}^{d+2n} \mid \exists \theta \in \mathbb{T}^{d+n}_{\rho_0} \text{ such that } |z - K_0(\theta)| < r \right\} , \tag{2.27}$$

for some $r > 0$ *and such that* $|V_\lambda|_{C^2, \mathcal{B}_r}$ *is finite.*

Define the error function as

$$e_{0,\theta} := V_{\lambda, K_0(\theta)} - \partial_\omega K_{0,\theta} \ .$$

Then there exists a constant $c > 0$ *depending on* d, n, σ, ρ_0, $\|DK_0\|_{\rho_0}$, r, $|V_\lambda|_{C^2,\mathcal{B}_r}$, $\left\| \frac{\partial V_\lambda}{\partial \lambda}\Big|_{\lambda=\lambda_0} \circ K_0 \right\|_{\rho_0}$*, and* $\left| \{\operatorname{avg}(\Lambda_0)\}^{-1} \right|$*, such that if*

$$0 < \delta_0 < \max\left\{1, \frac{\rho_0}{12}\right\}$$

and the error e_0 *satisfies the condition*

$$\|e_0\|_{\rho_0} \leq \min\left\{\gamma^4 \delta_0^{4\sigma},\, cr\gamma^2\delta_0^{2\sigma}\, \|e_0\|_{\rho_0}\right\} \ ,$$

then there exists a mapping $K \in \mathcal{W}_{\rho_0 - 6\delta_0}$ *and a vector* $\bar{\lambda} \in \mathbb{R}^{d+2n}$ *such that*

$$V_{\bar{\lambda},K(\theta)} = \partial_\omega K_\theta \ .$$

Moreover, the following inequalities are satisfied:

$$\begin{aligned} \|K - K_0\|_{\rho_0 - 6\delta_0} &< \frac{1}{c}\,\gamma^2\delta_0^{-2\sigma}\,\|e_0\|_{\rho_0} \\ \left|\bar{\lambda} - \lambda_0\right| &< \frac{1}{c}\,\gamma^2\delta_0^{-2\sigma}\,\|e_0\|_{\rho_0} \ . \end{aligned}$$

Chapter 3

True Solutions

In this chapter we will introduce the concept of an invariant torus – i.e., a true solution of the problem – and will prove some results for invariant tori. We will discuss in detail the geometry of an invariant torus, will develop some geometric ideas, and perform some calculations that will be useful for the construction of invariant tori in the following chapters.

In particular, we will introduce a special basis in the tangent space to the presymplectic manifold near the invariant torus that will utilize the geometry and the dynamics of the problem. This basis will be an important tool because the equations that need to be solved in order to find the invariant torus have a simpler form in this basis.

3.1 Invariant tori

We start with the definition of an invariant torus and then discuss the equations that will need to be solved in order to find an invariant torus.

Definition 3.1. *Let $V_\lambda \in \mathfrak{X}(\mathcal{P})$ be a $(d+2n)$-parameter family of presymplectic* *vector fields on the exact presymplectic manifold $\mathcal{P}$. If for some value $\bar{\lambda}$ of the parameter λ there exists an embedding*

$$K\colon \mathbb{T}^{d+n} \to \mathcal{P} \ ,$$

such that

$$V_{\bar{\lambda},K(\theta)} = \partial_\omega K_\theta := \sum_{\alpha=1}^{d+n} \frac{\partial K_\theta}{\partial \theta^\alpha}\, \omega^\alpha \tag{3.1}$$

and $\omega \in \mathbb{R}^{d+n}$ is a Diophantine vector, we call K an ***invariant torus*** *or a* ***true solution****.*

In (3.1), we think of the vector $\omega \in \mathbb{R}^{d+n}$ and the family $\{V_\lambda\}$ of presymplectic vector fields as given, while the special value $\bar{\lambda}$ of the parameter λ, as well as K are unknown. We may also refer to the image $\mathcal{K}$ (2.18) of $\mathbb{T}^{d+n}$ under K as an invariant torus.

In this chapter, we are concerned only with true solutions, but it seems appropriate to define the notion of approximate solution immediately after we have defined a true solution; we provide the formal definition in Chapter 4. Notice that we can write (3.1) as

$$V_{\bar{\lambda},K(\theta)} - \partial_\omega K_\theta = 0 \ .$$

If the difference $V_{\bar{\lambda}\,,\,K(\theta)} - \partial_\omega K_\theta$ is non-zero but is small in some norm, then we will call such K an **approximate solution**.

If K satisfies (3.1), then for the flow Φ_t of the vector field $V_{\bar{\lambda}}$, we have that

$$\Phi_t(K_\theta) = K_{\phi_t(\theta)} = K_{\theta+t\omega} \in \mathcal{K} \ , \tag{3.2}$$

thus the embedded torus $\mathcal{K}$ (2.18) is invariant under the flow of $V_{\bar{\lambda}}$. This is the reason for calling a solution of (3.1) an invariant torus. However, this terminology is somewhat of a misnomer. We are actually requiring more than being merely invariant; we are requiring that the motion on $\mathcal{K}$ be quasi-periodic and ω be Diophantine (Definition 2.9). We will call such a solution a **KAM torus**. Before we give the definition of quasi-periodic, first recall what it means for a vector to be independent over the rationals.

Definition 3.2. *A vector $\omega \in \mathbb{R}^m$ with $\omega = (\omega_1, \ldots, \omega_m)$ is said to be **independent over the rationals** if for $k_i \in \mathbb{Q}$, such that $\omega_1 k_1 + \cdots + \omega_m k_m = 0$, then $k_i = 0$ for all i.*

Definition 3.3. *The motion on an invariant torus $\mathcal{K}$ is said to be **quasi-periodic** if the dymanics can be conjugated to to a linear flow on a torus with frequency vector ω that is independent over the rationals.*

Remark 1. As noted in [56], the existence of a quasi-periodic solution leads to the existence of an embedded torus that is invariant under the action of Φ_t. However if we have an embedded torus that is invariant under the action of Φ_t, it need not come from a quasi-periodic solution because the motion could be different from an irrational rotation. In this dissertation, when we say invariant torus, we really mean the image of a quasi-periodic solution with Diophantine frequencies.

Remark 2. All Diophantine vectors are independent over the rationals, but the converse is not true.

3.2 Invariant tori are isotropic

Notational convention for this chapter. *In the rest of this chapter we will* *work with true solutions, which exist only for the value $\bar{\lambda}$ of the parameter λ in the family of presymplectic vector fields $\{V_\lambda\}$. Since in this chapter the parameter λ will always be equal to $\bar{\lambda}$, we will not write the subscript λ in the notation of the vector field, i.e., we set*

$$V := V_{\bar{\lambda}} \ .$$

We start this section with a definition.

Definition 3.4. *An invariant (in the sense of Definition 3.1) torus, $\mathcal{K} = K(\mathbb{T}^{d+n})$, in the exact presymplectic manifold $\mathcal{P}$ is said to be **isotropic** if the pull-back, $K^*\Omega \in \Omega^2(\mathbb{T}^{d+n})$, of the presymplectic form $\Omega \in \Omega^2(\mathcal{P})$ to the torus $\mathbb{T}^{d+n}$ vanishes identically.*

In Lemma 3.5 below we will prove that an invariant torus is isotropic. Similar results for maps are well-known for the case of submanifolds invariant with respect to symplectic or presymplectic maps (see, e.g., [26, Section 4, Lemma 1] or [4, Lemma 2.5]; our proof follows the same idea). The fact that the invariant torus is isotropic is crucial in the proof of Lemma 4.2 which, in turn, is essential for the bounds we will need to solve the linearized equation in Section 4.3.

Similarly to the linear operators J_p and $\widetilde{J}_q$ (introduced in (2.7) and (2.8)) that represents the presymplectic form Ω on $\mathcal{P}$ and symplectic form $\widetilde{\Omega}$ on $\mathcal{Q}$, we introduce the linear operator

$$L_\theta : T_\theta\mathbb{T}^{d+n} \to T_\theta\mathbb{T}^{d+n}$$

as the matrix representation of the pull-back $(K^*\Omega)_\theta$:

$$\langle \eta_\theta, L_\theta\, \zeta_\theta \rangle_{\mathbb{R}^{d+n}} = (K^*\Omega)_\theta\, (\eta_\theta, \zeta_\theta)\ , \qquad \eta_\theta, \zeta_\theta \in T_\theta \mathbb{T}^{d+n}\ . \tag{3.3}$$

We have

$$\begin{aligned}
(K^*\Omega)_\theta\, (\eta_\theta, \zeta_\theta) &= \Omega_{K(\theta)}\, (K_{*\theta}\, \eta_\theta, K_{*\theta}\, \zeta_\theta) \\
&= \left\langle DK_\theta\, \eta_\theta, J_{K(\theta)} DK_\theta\, \zeta_\theta \right\rangle_{\mathbb{R}^{d+n}} \\
&= \left(DK_\theta\, \eta_\theta\right)^\top J_{K(\theta)} DK_\theta\, \zeta_\theta \\
&= \eta_\theta^\top\, DK_\theta^\top J_{K(\theta)} DK_\theta\, \zeta_\theta \\
&= \langle \eta_\theta, DK_\theta^\top J_{K(\theta)} DK_\theta\, \zeta_\theta \rangle_{\mathbb{R}^{d+n}}\ ,
\end{aligned}$$

which yields the following explicit expression for the matrix elements of L_θ:

$$L_\theta = DK_\theta^\top\, J_{K(\theta)}\, DK_\theta \in \mathrm{M}_{d+n,d+n}(\mathbb{R})\ . \tag{3.4}$$

Lemma 3.5. *Let $\mathcal{P}$ be an exact presymplectic manifold with presymplectic form $\Omega \in \Omega^2(\mathcal{P})$, $V \in \mathfrak{X}(\mathcal{P})$ be a presymplectic vector field on $\mathcal{P}$, and let $K : \mathbb{T}^{d+n} \to \mathcal{P}$ be a true solution. Then the pull-back $K^*\Omega$ of the exact presymplectic form Ω to $\mathbb{T}^{d+n}$ and, hence, its matrix representation L_θ, vanish identically. This implies that the invariant torus $\mathcal{K}$* (2.18) *is an isotropic submanifold of $\mathcal{P}$.*

Proof. We will prove the lemma in two steps: first we will use the exactness of Ω to show that the average (over $\mathbb{T}^{d+n}$) of each matrix element of L is zero, and then will use the ergodicity of the flow $\theta \mapsto \theta + t\omega$ on $\mathbb{T}^{d+n}$ to demonstrate that $K^*\Omega$ and, therefore, L, are constant on $\mathbb{T}^{d+n}$.

Since the presymplectic form Ω is exact, there exists a 1-form $\tau \in \Omega^1(\mathcal{P})$ such that

$$\Omega = \mathrm{d}\tau \ .$$

Due to the commutativity of the exterior derivative with pull-backs,

$$K^*\Omega = K^*\left(\mathrm{d}\tau\right) = \mathrm{d}\left(K^*\tau\right) \ . \tag{3.5}$$

If

$$\tau_{K(\theta)} = \sum_{A=1}^{d+2n} \tau_A(K(\theta))\,\mathrm{d}x^A \ ,$$

then the pull-back $K^*\tau \in \Omega^1(\mathbb{T}^{d+n})$ is given by

$$(K^*\tau)_\theta = \sum_{\alpha=1}^{d+n} C_\alpha(\theta)\,\mathrm{d}\theta^\alpha \ ,$$

where

$$C_\alpha(\theta) = \sum_{A=1}^{d+2n} \tau_A(K(\theta))\,\frac{\partial K^A}{\partial \theta^\alpha}(\theta) \ .$$

The matrix representation of the pull-back $(K^*\Omega)_\theta$ is then

$$\left(L_\theta\right)^\alpha{}_\beta = \frac{\partial C_\alpha}{\partial \theta^\beta}(\theta) - \frac{\partial C_\beta}{\partial \theta^\alpha}(\theta) \ .$$

Because of the periodicity of the functions $C_\alpha : \mathbb{T}^{d+n} \to \mathbb{R}$,

$$\begin{aligned} \operatorname{avg}\left(\frac{\partial C_\alpha}{\partial \theta^\beta}\right) &= \int_{\mathbb{T}^{d+n}} \frac{\partial C_\alpha}{\partial \theta^\beta}(\theta)\,\mathrm{d}\theta^1\,\mathrm{d}\theta^2\,\cdots\,\mathrm{d}\theta^{d+n} \\ &= \int_{\mathbb{T}^{d+n-1}} \left(\int_{\mathbb{T}^1} \frac{\partial C_\alpha}{\partial \theta^\beta}(\theta)\,\mathrm{d}\theta^\beta\right)\,\mathrm{d}\theta^1\,\mathrm{d}\theta^2\,\cdots\,\mathrm{d}\theta^{\beta-1}\,\mathrm{d}\theta^{\beta+1}\,\mathrm{d}\theta^{d+n} \qquad (3.6) \\ &= 0 \ , \end{aligned}$$

which shows that the average of each matrix element of L_θ is identically zero on $\mathbb{T}^{d+n}$, so that

$$\operatorname{avg}(L) = 0 \ , \qquad \operatorname{avg}(K^*\Omega) = 0 \ . \tag{3.7}$$

Now we will prove that L and $K^*\Omega$ are constant on $\mathbb{T}^{d+n}$. Restrict the target space of the map K (2.17) from $\mathcal{P}$ to the image $\mathcal{K}$ of K, to obtain the diffeomorphism

$$K|_{\mathbb{T}^{d+n}\to\mathcal{K}} : \mathbb{T}^{d+n} \to \mathcal{K} \ . \tag{3.8}$$

Since the manifold $\mathcal{K}$ is invariant with respect to the flow of the vector field $V \in \mathfrak{X}(\mathcal{P})$, at the points of $\mathcal{K}$ the vector field is tangent to $\mathcal{K}$. Therefore, the restriction of the vector field V to $\mathcal{K}$ can be considered as a section of the tangent bundle of $\mathcal{K}$; let us denote this new vector field by $V|_{\mathcal{K}} \in \mathfrak{X}(\mathcal{K})$. Because of the same reasons, the Lie derivative with respect to V has a natural restriction to sections of any tensor power of the tangent and cotangent bundles of $\mathcal{K}$. The pull-back of $V|_{\mathcal{K}}$ by the diffeomorphism (3.8) is

$$K^*V := \left(K|_{\mathbb{T}^{d+n}\to\mathcal{K}}\right)_*^{-1} \left(V|_{\mathcal{K}}\right) \in \mathfrak{X}(\mathbb{R}^{d+n}) \ .$$

If we consider the constant $\omega \in \mathbb{R}^{d+n}$ as a tangent vector $\omega_\theta \in T_\theta(\mathbb{T}^{d+n})$, then the pull-back of $V_{K(\theta)} = K_{*\theta}\,\omega_\theta \in T_{K(\theta)}\mathcal{K}$ is

$$\begin{aligned}
(K^*V)_\theta &= \left[\left(K|_{\mathbb{T}^{d+n}\to\mathcal{K}}\right)^{-1}\right]_{*K(\theta)} V_{K(\theta)} \\
&= \left[\left(K|_{\mathbb{T}^{d+n}\to\mathcal{K}}\right)^{-1}\right]_{*K(\theta)} K_{*\theta}\,\omega_\theta \\
&= \left[\left(K|_{\mathbb{T}^{d+n}\to\mathcal{K}}\right)^{-1} \circ K\right]_{*\theta} \omega_\theta \\
&= \omega_\theta \in T_\theta(\mathbb{T}^{d+n}) \ .
\end{aligned}$$

Then the well-known property of the Lie derivative [2, Proposition 2.2.19]

$$K^* \mathcal{L}_V \Omega = \mathcal{L}_{K^* V} K^* \Omega$$

becomes

$$K^* \mathcal{L}_V \Omega = \mathcal{L}_\omega K^* \Omega$$

(in the last two equations, it is understood that all objects and operations were restricted to $\mathcal{K}$). Since the vector field V is presymplectic, $\mathcal{L}_V \Omega = 0$, which implies that the pull-back $K^*\Omega \in \Omega^2(\mathbb{T}^{d+n})$ of the presymplectic form to the torus $\mathbb{T}^{d+n}$ is constant on the orbits of the flow $\theta \mapsto \theta + t\omega$, $t \in \mathbb{R}$. But since ω is Diophantine, this flow is ergodic on $\mathbb{T}^{d+n}$, therefore $K^*\Omega$ is constant on $\mathbb{T}^{d+n}$:

$$K^*\Omega = \text{const} , \qquad L = \text{const} . \tag{3.9}$$

Putting together (3.7) and (3.9), we obtain the desired result. □

3.3 Construction and properties of an adapted basis of $T_{K(\theta)}\mathcal{P}$

In this section we will construct a special basis of $T_{K(\theta)}\mathcal{P}$ at any point $K(\theta) \in \mathcal{K}$ of the invariant torus. Our construction will utilize the geometric structure of the problem.

3.3.1 Adapted coordinates in $\mathbb{T}^{d+n}$ and a basis of $T_{K(\theta)}\mathcal{K}$

We start the construction by recalling that, by Definition 3.1, the tangent bundle $T\mathcal{K}$ of the invariant torus $\mathcal{K}$ is the push-forward of the tangent bundle $T\mathbb{T}^{d+n}$ of the torus $\mathbb{T}^{d+n}$ under the derivative K_* of the embedding K:

$$T_{K(\theta)}\mathcal{K} = K_{*\theta}\left(T_\theta \mathbb{T}^{d+n}\right) \ .$$

Therefore, every vector in $T_{K(\theta)}\mathcal{K}$ has the form $K_{*\theta}\,\eta_\theta$ for some $\eta_\theta \in T_\theta\mathbb{T}^{d+n}$. In matrix notations, $K_{*\theta}\,\eta_\theta$ is written as $DK_\theta\,\eta_\theta$, where the components, η_θ^α, of a vector

$$\eta_\theta = \sum_{\alpha=1}^{d+n} \left(\frac{\partial}{\partial\theta^\alpha}\right)_\theta \eta_\theta^\alpha$$

are written as a column vector of dimension $d+n$:

$$\eta_\theta = \left[\eta_\theta^1 \ \eta_\theta^2 \ \cdots \ \eta_\theta^{d+n}\right]^\top \ . \tag{3.10}$$

The vectors $\left\{\left(\frac{\partial}{\partial\theta^1}\right)_\theta, \dots, \left(\frac{\partial}{\partial\theta^{d+n}}\right)_\theta\right\}$ form a basis of $T_\theta\mathbb{T}^{d+n}$, hence the vectors

$$\left\{K_{*\theta}\left(\frac{\partial}{\partial\theta^1}\right)_\theta, \dots, K_{*\theta}\left(\frac{\partial}{\partial\theta^{d+n}}\right)_\theta\right\}$$

form a basis of $T_{K(\theta)}\mathcal{K}$. We have

$$\begin{aligned} K_{*\theta}\,\eta_\theta &= \sum_{\alpha=1}^{d+n} K_{*\theta}\left(\frac{\partial}{\partial\theta^\alpha}\right)_\theta \eta_\theta^\alpha \\ &= \sum_{\alpha=1}^{d+n}\left\{\sum_{A=1}^{d+2n}\left(\frac{\partial}{\partial x^A}\right)_{K(\theta)}\frac{\partial K^A}{\partial\theta^\alpha}(\theta)\right\}\eta_\theta^\alpha \\ &= \sum_{A=1}^{d+2n}\left(\frac{\partial}{\partial x^A}\right)_{K(\theta)}\left(\sum_{\alpha=1}^{d+n}\frac{\partial K^A}{\partial\theta^\alpha}(\theta)\,\eta_\theta^\alpha\right) \\ &= \sum_{A=1}^{d+2n}\left(\frac{\partial}{\partial x^A}\right)_{K(\theta)}(K_{*\theta}\,\eta_\theta)^A\ . \end{aligned} \tag{3.11}$$

If we write η_θ as a column vector of dimension $(d+n)$ as in (3.10), and $K_{*\theta}\eta_\theta$ as a column vector of dimension $(d+2n)$,

$$K_{*\theta}\eta_\theta = \left[(K_{*\theta}\eta_\theta)^1\ (K_{*\theta}\eta_\theta)^2\ \cdots\ (K_{*\theta}\eta_\theta)^{d+2n}\right]^\top ,$$

then (3.11) implies that

$$(K_{*\theta}\,\eta_\theta)^A = \sum_{\alpha=1}^{d+n}\frac{\partial K^A}{\partial\theta^\alpha}(\theta)\,\eta_\theta^\alpha\ ,$$

which can be written in the form

$$\begin{bmatrix} (K_{*\theta}\,\eta_\theta)^1 \\ (K_{*\theta}\,\eta_\theta)^2 \\ \vdots \\ (K_{*\theta}\,\eta_\theta)^{d+2n} \end{bmatrix} = \sum_{\alpha=1}^{d+n}\begin{bmatrix} \frac{\partial K_\theta^1}{\partial\theta^\alpha} \\ \frac{\partial K_\theta^2}{\partial\theta^\alpha} \\ \vdots \\ \frac{\partial K_\theta^{d+2n}}{\partial\theta^\alpha} \end{bmatrix}\eta_\theta^\alpha\ . \tag{3.12}$$

Therefore, we can think of the column vectors in the right-hand side of (3.12) as vectors in $T_{K(\theta)}\mathcal{K}$, and since for each $\eta_\theta \in T_\theta \mathbb{T}^{d+n}$, $K_{*\theta}\,\eta_\theta$ can be expressed as a superposition as in (3.12), these column vectors form a basis of $T_{K(\theta)}\mathcal{K}$. Since the column vectors in the right-hand side of (3.12) are the $(d+n)$ columns of the matrix of the derivative DK_θ of the map K,

$$DK_\theta = \Big[(DK_\theta)^A{}_\alpha \Big] = \left[\frac{\partial K^A}{\partial \theta^\alpha}(\theta) \right] \in \mathrm{M}_{d+2n,d+n}(\mathbb{R}) \ , \tag{3.13}$$

we will use the matrix DK_θ in the construction of a special basis for $T_{K(\theta)}\mathcal{K}$.

Now we will extend this construction in order to adapt our basis to the kernel of the presymplectic form Ω. Recalling that, according to Lemma 3.5, $\mathcal{K}$ is an isotropic manifold of dimension $(d+n)$, we see that the integrable distribution $\ker\Omega$ restricted to $\mathcal{K}$ must be a subbundle of the tangent bundle to $\mathcal{K}$:

$$(\ker\Omega)\,|_{\mathcal{K}} \subseteq T\mathcal{K} \ ,$$

i.e.,

$$\ker\Omega_{K(\theta)} \subseteq T_{K(\theta)}\mathcal{K} \qquad \forall K(\theta) \in \mathcal{K} \ .$$

Now recall that, by our choice of $\mathcal{P}$ (2.2), the first d coordinates in $\mathcal{P} = \mathbb{T}^d \times T^*\mathbb{T}^n$ correspond to the kernel of the presymplectic form Ω.

These geometric considerations motivate the following construction. Reorder the coordinates θ^α, $\alpha = 1, \ldots, d+n$, in $\mathbb{T}^{d+n}$ in such a way that the rank of the $d \times d$ submatrix in the upper left corner of the matrix DK_θ has full rank (this is always possible because $DK_\theta \in \mathrm{M}_{d+2n,d+n}(\mathbb{R})$ is of full rank). Then necessarily the $n \times n$ submatrix in the lower right corner of DK_θ will also be of full rank. Since the first d coordinates in $\mathcal{P}$ are along the kernel of Ω, this choice of coordinates in $\mathbb{T}^{d+n}$ guarantees that the span of the last n columns of DK_θ, thought of as vectors in $T_{K(\theta)}\mathcal{K}$, span an n-dimensional subspace of $T_{K(\theta)}\mathcal{K}$ that is transversal to $\ker\Omega_{K(\theta)}$ in $T_{K(\theta)}\mathcal{K}$.

We now introduce notations that will reflect this construction. Let $\dfrac{\partial K_\theta^\bullet}{\partial\theta^\alpha}$ be the αth column of DK_θ:

$$\frac{\partial K_\theta^\bullet}{\partial\theta^\alpha} := \begin{bmatrix} \dfrac{\partial K_\theta^1}{\partial\theta^\alpha} \\ \dfrac{\partial K_\theta^2}{\partial\theta^\alpha} \\ \vdots \\ \dfrac{\partial K_\theta^{d+2n}}{\partial\theta^\alpha} \end{bmatrix} \in \mathrm{M}_{d+2n,1}(\mathbb{R}) \ , \qquad \alpha = 1,\dots,d+n \ .$$

The bullet ($\bullet$) stands for the whole allowed range of coordinates. Denote the first d such columns by $(Z_\theta)^\bullet{}_\mu$, and the remaining n columns by $(X_\theta)^\bullet{}_a$:

$$(Z_\theta)^\bullet{}_\mu := \frac{\partial K_\theta^\bullet}{\partial\theta^\mu} \ , \qquad \mu = 1,\dots,d \ ,$$
$$(X_\theta)^\bullet{}_a := \frac{\partial K_\theta^\bullet}{\partial\theta^{d+a}} \ , \qquad a = 1,\dots,n \ .$$

Because of the choice of the coordinates in $\mathbb{T}^{d+n}$, the n-dimensional subspace

$$\mathrm{span}\ \{(X_\theta)^\bullet{}_1,\dots,(X_\theta)^\bullet{}_n\} \subseteq T_{K(\theta)}\mathcal{K}$$

is transversal in $T_{K(\theta)}\mathcal{K}$ both to the span of the vectors $(Z_\theta)^\bullet{}_1,\dots,(Z_\theta)^\bullet{}_d$ (which is d-dimensional) and to the d-dimensional subspace $\ker\Omega_{K(\theta)}$, i.e.,

$$\begin{aligned} &\mathrm{span}\ \{(X_\theta)^\bullet{}_1,\dots,(X_\theta)^\bullet{}_n\} \oplus \mathrm{span}\ \{(Z_\theta)^\bullet{}_1,\dots,(Z_\theta)^\bullet{}_d\} = T_{K(\theta)}\mathcal{K} \ , \\ &\mathrm{span}\ \{(X_\theta)^\bullet{}_1,\dots,(X_\theta)^\bullet{}_n\} \cap \mathrm{span}\ \{(Z_\theta)^\bullet{}_1,\dots,(Z_\theta)^\bullet{}_d\} = \{0\} \ , \end{aligned} \tag{3.14}$$

and

$$\text{span}\ \{(X_\theta)^\bullet{}_1, \ldots, (X_\theta)^\bullet{}_n\} \oplus \ker \Omega_{K(\theta)} = T_{K(\theta)}\mathcal{K}\ ,$$

$$\text{span}\ \{(X_\theta)^\bullet{}_1, \ldots, (X_\theta)^\bullet{}_n\} \cap \ker \Omega_{K(\theta)} = \{0\}\ .$$

We stack the vectors $(Z_\theta)^\bullet{}_1, \ldots, (Z_\theta)^\bullet{}_d$ together to form the matrix

$$Z_\theta := [(Z_\theta)^\bullet{}_1\ \ldots\ (Z_\theta)^\bullet{}_d] \in \mathrm{M}_{d+2n,d}(\mathbb{R})\ , \tag{3.15}$$

and do the same with the vectors $(X_\theta)^\bullet{}_1, \ldots, (X_\theta)^\bullet{}_n$ to construct the matrix

$$X_\theta := [(X_\theta)^\bullet{}_1\ \ldots\ (X_\theta)^\bullet{}_n] \in \mathrm{M}_{d+2n,n}(\mathbb{R})\ . \tag{3.16}$$

With these notations, the derivative DK_θ (3.13) of the map K can be written as

$$DK_\theta = [\, Z_\theta\ X_\theta \,] = [\, (Z_\theta)^\bullet{}_1\ \cdots\ (Z_\theta)^\bullet{}_d\ (X_\theta)^\bullet{}_1\ \cdots\ (X_\theta)^\bullet{}_n \,]\ . \tag{3.17}$$

Now we will also introduce convenient notations for the coordinates $x = \left(x^A\right)$, $A = 1, \ldots, d+2n$, in $\mathcal{P}$ (2.2). We denote the first d coordinates by putting an underscore, and the rest of the coordinates by putting a tilde:

$$x = \left(x^A\right) = (\underline{x}, \widetilde{x}) = \left(\underline{x}^\mu, \widetilde{x}^i\right)\ , \qquad \mu = 1, \ldots, d\,, \quad i = 1, \ldots, 2n\,.$$

In more detail,

$$\underline{x} = (\underline{x}^\mu) = \left(\underline{x}^1, \ldots, \underline{x}^d\right) = \left(x^1, \ldots x^d\right)\ ,$$

$$\widetilde{x} = (\tilde{x}^i) = \left(\widetilde{x}^1, \ldots, \widetilde{x}^{2n}\right) = \left(x^{d+1}, \ldots x^{d+2n}\right)\ .$$

This notation is compatible with the notation introduced in (2.24), but now its meaning is more clear.

We will also widely use the underscore and tilde notations in matrices with $d+2n$ rows – the underscore for the first d rows, and the tilde for the remaining $2n$ rows. In particular, the derivative of K (3.17) will be written as

$$DK_\theta = [Z_\theta \ X_\theta] = \begin{bmatrix} \underline{Z}_\theta & \underline{X}_\theta \\ \widetilde{Z}_\theta & \widetilde{X}_\theta \end{bmatrix} , \tag{3.18}$$

where

$$\begin{aligned} &\underline{Z}_\theta \in \mathrm{M}_{d,d}(\mathbb{R}) \ , \qquad &&\underline{X}_\theta \in \mathrm{M}_{d,n}(\mathbb{R}) \ , \\ &\widetilde{Z}_\theta \in \mathrm{M}_{2n,d}(\mathbb{R}) \ , \qquad &&\widetilde{X}_\theta \in \mathrm{M}_{2n,n}(\mathbb{R}) \ . \end{aligned} \tag{3.19}$$

3.3.2 Adapted basis of $T_{K(\theta)}\mathcal{P}$

To construct a geometrically natural basis of $T_{K(\theta)}\mathcal{P}$, we need n more vectors that are linearly independent from $(Z_\theta)^\bullet{}_\mu$ and $(X_\theta)^\bullet{}_a$ and should span the complement of $T_{K(\theta)}\mathcal{K}$ in $T_{K(\theta)}\mathcal{P}$. Since these new vectors, together with the n vectors $(X_\theta)^\bullet{}_a$, should form a basis of the tangent space $T_{K(\theta)}\mathcal{Q}$ to the symplectic manifold $\mathcal{Q}$ (2.6), it is natural to use the symplectic form $\widetilde{\Omega}$ on $\mathcal{Q}$, whose matrix representation is given by the antisymmetric non-degenerate matrix $\widetilde{J}_{K(\theta)}$ defined in (2.8). Before constructing these new vectors, we emphasize that, strictly speaking, instead of $\widetilde{J}_{K(\theta)}$ we should write $\widetilde{J}_{\pi^{\mathcal{Q}}(K(\theta))}$, where $\pi^{\mathcal{Q}}$ (2.4) is the projection from the presymplectic manifold $\mathcal{P}$ onto the (abstract) symplectic manifold $\mathcal{Q}$. However, we will use the shorter notation $\widetilde{J}_{K(\theta)}$ – which emphasizes the fact that we work with concrete geometric objects rather than with abstract objects defined on factormanifolds.

Clearly, there are many ways to choose the new n vectors, but we will make a special choice such that they, together with the vectors $(X_\theta)^\bullet{}_a$, form a symplectic basis of $T\mathcal{Q}$ if we "mod out" the kernel of the presymplectic form. To achieve this, we will use the Gramian matrix of the vectors $(X_\theta)^\bullet{}_1, \ldots, (X_\theta)^\bullet{}_n$, i.e., the matrix of the inner products of these vectors with respect to the Euclidean inner product on $T^*\mathbb{T}^n$. The Gramian of the vectors $(X_\theta)^\bullet{}_1, \ldots, (X_\theta)^\bullet{}_n$ can be written as the matrix $X_\theta^\top X_\theta \in \mathrm{M}_{n,n}(\mathbb{R})$. Since the vectors $(X_\theta)^\bullet{}_1, \ldots, (X_\theta)^\bullet{}_n$ are linearly independent, the rank of their Gramian is n. This allows us to define the (symmetric) matrix

$$R_\theta := \left(\widetilde{X}_\theta^\top \widetilde{X}_\theta\right)^{-1} \in \mathrm{M}_{n,n}(\mathbb{R}) \ . \tag{3.20}$$

Define the matrices

$$\widetilde{Y}_\theta := \widetilde{J}_{K(\theta)}^{-1}\, \widetilde{X}_\theta\, R_\theta \in \mathrm{M}_{2n,n}(\mathbb{R}) \ , \tag{3.21}$$

and

$$Y_\theta := \begin{bmatrix} 0 \\ \\ \widetilde{Y}_\theta \end{bmatrix} = \begin{bmatrix} 0 \\ \\ \widetilde{J}_{K(\theta)}^{-1}\, \widetilde{X}_\theta\, R_\theta \end{bmatrix} \in \mathrm{M}_{d+2n,n}(\mathbb{R}) \ ; \tag{3.22}$$

in our notations for the components, $\underline{Y}_\theta = 0$ (recall (3.18)). Since the matrices $\widetilde{J}_{K(\theta)}^{-1}$, $\widetilde{X}_\theta$, and R_θ are all of rank n, $\widetilde{Y}_\theta$ is also of maximal rank:

$$\operatorname{rank} \widetilde{Y}_\theta = n \ .$$

Similarly to (3.15) and (3.16), we think of the n columns of Y_θ as column vectors of dimension $(d + 2n)$:

$$Y_\theta =: \left[(Y_\theta)^\bullet{}_1 \ \cdots \ (Y_\theta)^\bullet{}_n\right] \ ,$$

where

$$(Y_\theta)^\bullet{}_a = \left[(Y_\theta)^1{}_a \; (Y_\theta)^2{}_a \; \cdots \; (Y_\theta)^{d+2n}{}_a\right]^\top \in \mathrm{M}_{d+2n,1}(\mathbb{R}) \;, \qquad a = 1, \ldots, n \;.$$

We think of $(Y_\theta)^\bullet{}_a$ as vectors in $T_{K(\theta)}\mathcal{P}$. With the definitions (2.7), (2.8), (2.9), (3.20), (3.21), and (3.22), we obtain

$$\begin{aligned}
\Omega_{K(\theta)}\big((X_\theta)^\bullet{}_a, (Y_\theta)^\bullet{}_b\big) &= \big\langle (X_\theta)^\bullet{}_a, J_{K(\theta)}\,(Y_\theta)^\bullet{}_b \big\rangle_{\mathbb{R}^{d+2n}} \\
&= \left((X_\theta)^\bullet{}_a\right)^\top J_{K(\theta)}\,(Y_\theta)^\bullet{}_b \\
&= \left(\begin{bmatrix} \underline{X}_\theta \\ \widetilde{X}_\theta \end{bmatrix}^\bullet{}_a\right)^\top \begin{bmatrix} 0 & 0 \\ 0 & \widetilde{J}_{K(\theta)} \end{bmatrix} \begin{bmatrix} 0 \\ \widetilde{J}^{-1}_{K(\theta)}\,\widetilde{X}_\theta\,R_\theta \end{bmatrix}^\bullet{}_b \\
&= \left[\underline{X}_\theta^\top \; \widetilde{X}_\theta^\top\right]^a{}_\bullet \begin{bmatrix} 0 \\ \widetilde{X}_\theta\,R_\theta \end{bmatrix}^\bullet{}_b \\
&= (\widetilde{X}_\theta^\top \widetilde{X}_\theta R_\theta)_{ab} = (\mathbb{I}_n)_{ab} = \delta_{ab} \;,
\end{aligned} \tag{3.23}$$

where $\mathbb{I}_n$ is the unit $n \times n$ matrix, and δ_{ab} is Kronecker's symbol. Therefore, the vectors $(X_\theta)^\bullet{}_a$ and $(Y_\theta)^\bullet{}_b$ form a symplectic basis of $T_{K(\theta)}\mathcal{Q} \cong (T_{K(\theta)}\mathcal{P})/\ker\Omega_{K(\theta)}$. We will write (3.23) symbolically as

$$\Omega_{K(\theta)}\,(X_\theta, Y_\theta) = X_\theta^\top\,J_{K(\theta)}\,Y_\theta = \widetilde{X}_\theta^\top\,\widetilde{J}_{K(\theta)}\,\widetilde{Y}_\theta = \mathbb{I}_n \;. \tag{3.24}$$

The equality (3.23) implies that the vectors $(Y_\theta)^\bullet{}_a$ are linearly independent from the vectors $(Z_\theta)^\bullet{}_\mu$ and $(X_\theta)^\bullet{}_a$, so that all these $(d+2n)$ vectors form a basis of $T_{K(\theta)}\mathcal{P}$:

$$\operatorname{span}\left\{\left\{(Z_\theta)^\bullet{}_\mu\right\}_{\mu=1}^{d}, \left\{(X_\theta)^\bullet{}_a\right\}_{a=1}^{n}, \left\{(Y_\theta)^\bullet{}_a\right\}_{a=1}^{n}\right\} = T_{K(\theta)}\mathcal{P} \ . \tag{3.25}$$

The basis of $(T\mathcal{P})|_{\mathcal{K}}$ consisting of the vector fields $Z^\bullet{}_\mu$, $X^\bullet{}_a$, and $Y^\bullet{}_a$ is convenient for several reasons:

(a) the vector fields $Z^\bullet{}_\mu$ and $X^\bullet{}_a$ are a basis of the subbundle $T\mathcal{K} \subseteq T\mathcal{P}$ which is invariant with respect to the flow Φ_t of the vector field V, which implies that in the basis $Z^\bullet{}_\mu$, $X^\bullet{}_a$, $Y^\bullet{}_a$, the matrix of the transformation Φ_t will be upper block triangular, with the lower left $2n \times d$ block being zero;

(b) since the invariant torus $\mathcal{K}$ is an isotropic submanifold as we proved in Lemma 3.5, if the two arguments of $\Omega_{K(\theta)}$ are vectors from $T_{K(\theta)}\mathcal{K}$, the result is zero:

$$\Omega\big(Z^\bullet{}_\mu, Z^\bullet{}_\nu\big) = 0 \ ,$$

$$\Omega\big(Z^\bullet{}_\mu, X^\bullet{}_a\big) = 0 \ ,$$

$$\Omega\big(X^\bullet{}_a, X^\bullet{}_b\big) = 0 \ ,$$

for all $\mu, \nu = 1, \ldots, d$, and all $a, b = 1, \ldots, n$;

(c) the vector fields $\widetilde{X}^\bullet{}_a$ and $\widetilde{Y}^\bullet{}_b$ form a symplectic basis of $(T\mathcal{Q})|_{\mathcal{K}}$, with respect to the symplectic form $\widetilde{\Omega}$ (cf. (3.23)).

Below we summarize the properties of the basis of $T_{K(\theta)}\mathcal{P}$ constructed above, using matrix notations as in (3.24) (which will be more convenient for calculations):

$$
\begin{aligned}
Z_\theta^\top J_{K(\theta)} Z_\theta &= \widetilde{Z}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{Z}_\theta = 0 \ , \\
Z_\theta^\top J_{K(\theta)} X_\theta &= \widetilde{Z}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{X}_\theta = 0 \ , \\
Z_\theta^\top J_{K(\theta)} Y_\theta &= \widetilde{Z}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{Y}_\theta = \widetilde{Z}_\theta^\top \widetilde{X}_\theta R_\theta \ , \\
X_\theta^\top J_{K(\theta)} X_\theta &= \widetilde{X}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{X}_\theta = 0 \ , \\
X_\theta^\top J_{K(\theta)} Y_\theta &= \widetilde{X}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{Y}_\theta = \mathbb{I}_n \ , \\
Y_\theta^\top J_{K(\theta)} Y_\theta &= \widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{Y}_\theta = -R_\theta \widetilde{X}_\theta^\top \widetilde{J}_{K(\theta)}^{-1} \widetilde{X}_\theta R_\theta \ .
\end{aligned}
\tag{3.26}
$$

The fact that J and $\widetilde{J}$ are antisymmetric implies automatically that

$$
\begin{aligned}
X_\theta^\top J_{K(\theta)} Z_\theta &= \widetilde{X}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{Z}_\theta = 0 \ , \\
Y_\theta^\top J_{K(\theta)} Z_\theta &= \widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{Z}_\theta = -R_\theta \widetilde{X}_\theta^\top \widetilde{Z}_\theta \ , \\
Y_\theta^\top J_{K(\theta)} X_\theta &= \widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{X}_\theta = -\mathbb{I}_n \ .
\end{aligned}
$$

3.3.3 Presymplecticity of V at $\mathcal{K}$ in adapted coordinates

In this section we will rewrite the presymplecticity condition (2.3) (written in matrix form in (2.26)) in the adapted coordinates introduced above. In the adapted basis we write the vectors from $T_{K(\theta)}\mathcal{P}$, $K(\theta) \in \mathcal{K}$, in the form

$$
U_\theta = \begin{bmatrix} \underline{U}_\theta \\ \widetilde{U}_\theta \end{bmatrix} \in \mathrm{M}_{d+2n,1}(\mathbb{R}) \ ,
$$

with $\underline{U}_\theta \in \mathrm{M}_{d,1}(\mathbb{R})$, $\widetilde{U}_\theta \in \mathrm{M}_{2n,1}(\mathbb{R})$.

Let $V \in \mathfrak{X}(\mathcal{P})$ be a presymplectic vector field, and let $DV_{K(\theta)}$ be its derivative at a point $K(\theta) \in \mathcal{K}$. We introduce the notation

$$DV_{K(\theta)} =: \begin{bmatrix} \dfrac{\partial \underline{V}}{\partial \underline{x}} & \dfrac{\partial \underline{V}}{\partial \widetilde{x}} \\[2ex] \dfrac{\partial \widetilde{V}}{\partial \underline{x}} & \dfrac{\partial \widetilde{V}}{\partial \widetilde{x}} \end{bmatrix}_{K(\theta)}$$

where the subscript $K(\theta)$ means that all partial derivatives in the right-hand side are evaluated at the point $K(\theta) \in \mathcal{K}$, and

$$\left[\frac{\partial \underline{V}}{\partial \underline{x}}\right] = \left[\frac{\partial \underline{V}^\mu}{\partial \underline{x}^\nu}\right] = \left[\frac{\partial V^\mu}{\partial x^\nu}\right] \in \mathrm{M}_{d,d}(\mathbb{R}) \ , \qquad \mu, \nu = 1, \ldots, d \ ,$$

$$\left[\frac{\partial \underline{V}}{\partial \widetilde{x}}\right] = \left[\frac{\partial \underline{V}^\mu}{\partial \widetilde{x}^i}\right] = \left[\frac{\partial V^\mu}{\partial x^{d+i}}\right] \in \mathrm{M}_{d,2n}(\mathbb{R}) \ , \qquad \mu = 1, \ldots, d \ , \quad i = 1, \ldots, 2n \ ,$$

$$\left[\frac{\partial \widetilde{V}}{\partial \underline{x}}\right] = \left[\frac{\partial \widetilde{V}^i}{\partial \underline{x}^\mu}\right] = \left[\frac{\partial V^{d+i}}{\partial x^\mu}\right] \in \mathrm{M}_{2n,d}(\mathbb{R}) \ , \qquad i = 1, \ldots, 2n \ , \quad \mu = 1, \ldots, d \ ,$$

$$\left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right] = \left[\frac{\partial \widetilde{V}^i}{\partial \widetilde{x}^j}\right] = \left[\frac{\partial V^{d+i}}{\partial x^{d+j}}\right] \in \mathrm{M}_{d,d}(\mathbb{R}) \ , \qquad i, j = 1, \ldots, 2n \ .$$

Using these notations, we can write the presymplecticity condition (2.26) in adapted coordinates as

$$\begin{bmatrix} 0 & 0 \\[2ex] 0 & (D\widetilde{J})_{K(\theta)} V_{K(\theta)} \end{bmatrix} + \begin{bmatrix} \dfrac{\partial \underline{V}}{\partial \underline{x}} & \dfrac{\partial \underline{V}}{\partial \widetilde{x}} \\[2ex] \dfrac{\partial \widetilde{V}}{\partial \underline{x}} & \dfrac{\partial \widetilde{V}}{\partial \widetilde{x}} \end{bmatrix}_{K(\theta)}^{\top} \begin{bmatrix} 0 & 0 \\[2ex] 0 & \widetilde{J}_{K(\theta)} \end{bmatrix}$$
$$+ \begin{bmatrix} 0 & 0 \\[2ex] 0 & \widetilde{J}_{K(\theta)} \end{bmatrix} \begin{bmatrix} \dfrac{\partial \underline{V}}{\partial \underline{x}} & \dfrac{\partial \underline{V}}{\partial \widetilde{x}} \\[2ex] \dfrac{\partial \widetilde{V}}{\partial \underline{x}} & \dfrac{\partial \widetilde{V}}{\partial \widetilde{x}} \end{bmatrix}_{K(\theta)} = 0 \ ,$$

which can be rewritten as

$$\begin{bmatrix} 0 & \left[\dfrac{\partial \widetilde{V}}{\partial \underline{x}}\right]^\top \widetilde{J} \\ \widetilde{J}\left[\dfrac{\partial \widetilde{V}}{\partial \underline{x}}\right] & (D\widetilde{J})\,V + \left[\dfrac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]^\top \widetilde{J} + \widetilde{J}\left[\dfrac{\partial \widetilde{V}}{\partial \widetilde{x}}\right] \end{bmatrix}_{K(\theta)} = 0 \ . \tag{3.27}$$

The antisymmetry of $\widetilde{J}_{K(\theta)}$ implies that

$$\left[\frac{\partial \widetilde{V}}{\partial \underline{x}}\right]^\top_{K(\theta)} \widetilde{J}_{K(\theta)} = -\left(\widetilde{J}_{K(\theta)}\left[\frac{\partial \widetilde{V}}{\partial \underline{x}}\right]_{K(\theta)}\right)^\top ,$$

so that the off-diagonal entries of the matrix in (3.27) yield the condition

$$\widetilde{J}_{K(\theta)}\left[\frac{\partial \widetilde{V}}{\partial \underline{x}}\right]_{K(\theta)} = 0 \ .$$

Since $\widetilde{J}_{K(\theta)}$ is an invertible matrix, multiplying this identity on the right by $\widetilde{J}^{-1}_{K(\theta)}$, we obtain the following consequence of the presymplecticity of the vector field V:

$$\left[\frac{\partial \widetilde{V}}{\partial \underline{x}}\right]_{K(\theta)} = 0 \ .$$

The condition coming from the lower right corner of the matrix in (3.27) reads

$$(D\widetilde{J})_{K(\theta)} V_{K(\theta)} + \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]^\top_{K(\theta)} \widetilde{J}_{K(\theta)} + \widetilde{J}_{K(\theta)}\left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)} = 0 \ . \tag{3.28}$$

Note that the matrix $\left(\left[\dfrac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]^\top_{K(\theta)} \widetilde{J}_{K(\theta)} + \widetilde{J}_{K(\theta)}\left[\dfrac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}\right)$ is antisymmetric because it is the difference between the matrix $\widetilde{J}_{K(\theta)} \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}$ and its transposed, which in turn follows from the antisymmetry of $\widetilde{J}_{K(\theta)}$:

$$\left(\widetilde{J}_{K(\theta)} \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}\right)^\top = \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^\top \widetilde{J}_{K(\theta)}^\top = -\left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^\top \widetilde{J}_{K(\theta)} \ .$$

The antisymmetry of the matrix $\left(\left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^\top \widetilde{J}_p + \widetilde{J}_{K(\theta)} \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}\right)$ is consistent with the anti-symmetry of the term $(D\widetilde{J})_{K(\theta)} V_{K(\theta)}$ in (3.28).

To summarize, we obtained that the presymplecticity of the vector field V implies that the matrix $DV_{K(\theta)}$ is block upper triangular:

$$DV_{K(\theta)} = \begin{bmatrix} \frac{\partial \underline{V}}{\partial \underline{x}} & \frac{\partial \underline{V}}{\partial \widetilde{x}} \\ 0 & \frac{\partial \widetilde{V}}{\partial \widetilde{x}} \end{bmatrix}_{K(\theta)} , \tag{3.29}$$

and its component $\left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}$ satisfies (3.28).

3.4 Change of basis matrix M_θ: definition

The adapted basis $\{(Z_\theta)^\bullet{}_\mu\}_{\mu=1}^d$, $\{(X_\theta)^\bullet{}_a\}_{a=1}^n$, $\{(Y_\theta)^\bullet{}_a\}_{a=1}^n$ of $T_{K(\theta)}\mathcal{P}$ constructed in Section 3.3 has properties that are very useful for our analysis. Given an arbitrary column vector U_θ, considered as an element of $T_{K(\theta)}\mathcal{P}$, we can find its components in the adapted basis as follows. Define the **change of basis matrix**

M_θ of all vectors from the adapted basis, written as column vectors:

$$M_\theta := [DK_\theta \ \ Y_\theta] = [Z_\theta \ X_\theta \ Y_\theta] = \begin{bmatrix} \underline{Z}_\theta & \underline{X}_\theta & 0 \\ \widetilde{Z}_\theta & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} \in \mathrm{M}_{d+2n,d+2n}(\mathbb{R}) \ . \tag{3.30}$$

Then the vector U_θ can be written as a superposition of the vectors from the adapted basis as follows:

$$\begin{aligned} U_\theta &= M_\theta \, \xi_\theta \\ &= \sum_{\mu=1}^{d} (Z_\theta)^{\bullet}{}_{\mu} \, \xi_\theta^{\mu} + \sum_{a=1}^{n} (X_\theta)^{\bullet}{}_{a} \, \xi_\theta^{d+a} + \sum_{a=1}^{n} (Y_\theta)^{\bullet}{}_{a} \, \xi_\theta^{d+n+a} \ . \end{aligned} \tag{3.31}$$

In the adapted basis, if we write the $(d+2n)$ components of the vector ξ_θ as three blocks of length d, n, and n, as in the representation (3.31), then the vectors from $T_{K(\theta)}\mathcal{K}$ have the form $\xi_\theta = [* \ * \ 0]^\top$, where the stars represent numbers that are generally non-zero.

3.5 Change of basis matrix M_θ: computations

In this section we will perform some computations related to the change of basis matrix M_θ (3.30), which will be needed in Chapter 4.

Differentiating the invariance condition (3.1), we obtain

$$DV_{K(\theta)} \, DK_\theta = \partial_\omega \, DK_\theta \ . \tag{3.32}$$

This, together with the definition (3.30) of M_θ, gives us

$$(DV_{K(\theta)} - \partial_\omega) M_\theta = \begin{bmatrix} 0 & 0 & (DV_{K(\theta)} - \partial_\omega) Y_\theta \end{bmatrix} \ ,$$

Our first goal is to find an explicit expression for $(DV_{K(\theta)} - \partial_\omega)Y_\theta$. To this end we have to compute

$$(DV_{K(\theta)} - \partial_\omega)Y_\theta = \begin{bmatrix} \dfrac{\partial \underline{V}}{\partial \underline{x}} & \dfrac{\partial \underline{V}}{\partial \widetilde{x}} \\ 0 & \dfrac{\partial \widetilde{V}}{\partial \widetilde{x}} \end{bmatrix}_{K(\theta)} \begin{bmatrix} 0 \\ \widetilde{Y}_\theta \end{bmatrix} - \begin{bmatrix} 0 \\ \partial_\omega \widetilde{Y}_\theta \end{bmatrix}$$

$$= \begin{bmatrix} \left[\dfrac{\partial \underline{V}}{\partial \widetilde{x}}\right]_{K(\theta)} \widetilde{Y}_\theta \\ \left[\dfrac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)} \widetilde{Y}_\theta - \partial_\omega \widetilde{Y}_\theta \end{bmatrix} .$$

The computation of $(DV_{K(\theta)} - \partial_\omega)M_\theta$ is performed in Sections 3.5.1–3.5.2, and in Section 3.5.3 we represent this expression in a special form. In Section 3.5.4 we derive a factorization of M_θ and use it to compute M_θ^{-1}. In Section 3.5.5 we write down other factorizations of M_θ and M_θ^{-1} which will be useful in our analysis in the next chapter.

3.5.1 Computing $\partial_\omega \widetilde{Y}_\theta$

We break this calculation into several parts.

(a) By the Leibniz rule,

$$\begin{aligned} \partial_\omega \widetilde{Y}_\theta &= \partial_\omega \left(\widetilde{J}^{-1}_{K(\theta)} \widetilde{X}_\theta R_\theta \right) \\ &= \partial_\omega \big(\widetilde{J}^{-1}_{K(\theta)} \big) \widetilde{X}_\theta R_\theta + \widetilde{J}^{-1}_{K(\theta)} \partial_\omega \big(\widetilde{X}_\theta \big) R_\theta + \widetilde{J}^{-1}_{K(\theta)} \widetilde{X}_\theta \partial_\omega R_\theta \ . \end{aligned}$$

(b) Using the invariance condition (3.1) and the presymplecticity condition (3.28), we obtain

$$
\begin{aligned}
\partial_\omega\big(\widetilde{J}_{K(\theta)}\big) &= (D\widetilde{J})_{K(\theta)}\,\partial_\omega K_\theta \\
&= (D\widetilde{J})_{K(\theta)}\,V_{K(\theta)} \\
&= -\left(\widetilde{J}_{K(\theta)}\left[\frac{\partial\widetilde{V}}{\partial\widetilde{x}}\right]_{K(\theta)} + \left[\frac{\partial\widetilde{V}}{\partial\widetilde{x}}\right]^\top_{K(\theta)}\widetilde{J}_{K(\theta)}\right)\ .
\end{aligned}
$$

This, together with the elementary identity

$$
0 = \partial_\omega\big(\mathbb{I}_{2n}\big) = \partial_\omega\big(\widetilde{J}^{-1}_{K(\theta)}\widetilde{J}_{K(\theta)}\big) = \partial_\omega\big(\widetilde{J}^{-1}_{K(\theta)}\big)\widetilde{J}_{K(\theta)} + \widetilde{J}^{-1}_{K(\theta)}\partial_\omega\big(\widetilde{J}_{K(\theta)}\big)\,,
$$

yields

$$
\begin{aligned}
\partial_\omega\big(\widetilde{J}^{-1}_{K(\theta)}\big) &= -\widetilde{J}^{-1}_{K(\theta)}\partial_\omega\big(\widetilde{J}_{K(\theta)}\big)\widetilde{J}^{-1}_{K(\theta)} \\
&= \left[\frac{\partial\widetilde{V}}{\partial\widetilde{x}}\right]_{K(\theta)}\widetilde{J}^{-1}_{K(\theta)} + \widetilde{J}^{-1}_{K(\theta)}\left[\frac{\partial\widetilde{V}}{\partial\widetilde{x}}\right]^\top_{K(\theta)}\ .
\end{aligned}
\tag{3.33}
$$

(c) From the invariance identity (3.32), written in components as

$$
\begin{bmatrix} \dfrac{\partial\underline{V}}{\partial\underline{x}} & \dfrac{\partial\underline{V}}{\partial\widetilde{x}} \\ 0 & \dfrac{\partial\widetilde{V}}{\partial\widetilde{x}} \end{bmatrix}_{K(\theta)} \begin{bmatrix} \underline{Z}_\theta & \underline{X}_\theta \\ \widetilde{Z}_\theta & \widetilde{X}_\theta \end{bmatrix} = \begin{bmatrix} \partial_\omega\underline{Z}_\theta & \partial_\omega\underline{X}_\theta \\ \partial_\omega\widetilde{Z}_\theta & \partial_\omega\widetilde{X}_\theta \end{bmatrix}
$$

(recall (3.29)), we obtain

$$
\partial_\omega\widetilde{X}_\theta = \left[\frac{\partial\widetilde{V}}{\partial\widetilde{x}}\right]_{K(\theta)}\widetilde{X}_\theta\ . \tag{3.34}
$$

(d) From the definition (3.20) of R_θ and the expression for $\partial_\omega \widetilde{X}_\theta$ we have

$$
\begin{aligned}
0 &= \partial_\omega \big(\mathbb{I}_{2n}\big) \\
&= \partial_\omega \left(R_\theta \widetilde{X}_\theta^\top \widetilde{X}_\theta\right) \\
&= \partial_\omega \big(R_\theta\big) \widetilde{X}_\theta^\top \widetilde{X}_\theta + R_\theta \partial_\omega \big(\widetilde{X}_\theta^\top\big) \widetilde{X}_\theta + R_\theta \widetilde{X}_\theta^\top \partial_\omega \widetilde{X}_\theta \\
&= \partial_\omega \big(R_\theta\big) R_\theta^{-1} + R_\theta \big(\partial_\omega \widetilde{X}_\theta\big)^\top \widetilde{X}_\theta + R_\theta \widetilde{X}_\theta^\top \partial_\omega \widetilde{X}_\theta \\
&= \partial_\omega \big(R_\theta\big) R_\theta^{-1} + 2 R_\theta \widetilde{X}_\theta^\top \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^{\mathrm{sym}} \widetilde{X}_\theta \ ,
\end{aligned}
$$

where $\left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^{\mathrm{sym}}$ stands for the symmetrization of the matrix $\left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}$:

$$
\left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^{\mathrm{sym}} := \frac{1}{2}\left(\left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)} + \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^\top\right) \ . \tag{3.35}
$$

This yields

$$
\partial_\omega R_\theta = -2 R_\theta \widetilde{X}_\theta^\top \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^{\mathrm{sym}} \widetilde{X}_\theta R_\theta \ . \tag{3.36}
$$

(e) Putting everything together:

$$
\begin{aligned}
\partial_\omega \widetilde{Y}_\theta &= \partial_\omega \left(\widetilde{J}^{-1}_{K(\theta)} \widetilde{X}_\theta R_\theta \right) \\
&= \partial_\omega \big(\widetilde{J}^{-1}_{K(\theta)} \big) \widetilde{X}_\theta R_\theta + \widetilde{J}^{-1}_{K(\theta)} \partial_\omega \big(\widetilde{X}_\theta \big) R_\theta + \widetilde{J}^{-1}_{K(\theta)} \widetilde{X}_\theta \partial_\omega R_\theta \\
&= \left(\left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}} \right]_{K(\theta)} \widetilde{J}^{-1}_{K(\theta)} + \widetilde{J}^{-1}_{K(\theta)} \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}} \right]^{\top}_{K(\theta)} \right) \widetilde{X}_\theta R_\theta \\
&\quad + \widetilde{J}^{-1}_{K(\theta)} \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}} \right]_{K(\theta)} \widetilde{X}_\theta R_\theta \\
&\quad - 2 \widetilde{J}^{-1}_{K(\theta)} \widetilde{X}_\theta R_\theta \widetilde{X}_\theta^{\top} \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}} \right]^{\mathrm{sym}}_{K(\theta)} \widetilde{X}_\theta R_\theta \ .
\end{aligned}
$$

Recalling the definition (3.21) of $\widetilde{Y}_\theta$ and introducing the operator

$$
\widetilde{\Pi}_\theta := \mathbb{I}_{2n} - \widetilde{X}_\theta R_\theta \widetilde{X}_\theta^{\top} : \mathbb{R}^{2n} \to \mathbb{R}^{2n} \ , \tag{3.37}
$$

we rewrite this as

$$
\partial_\omega \widetilde{Y}_\theta = \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}} \right]_{K(\theta)} \widetilde{Y}_\theta + 2 \widetilde{J}^{-1}_{K(\theta)} \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}} \right]^{\mathrm{sym}}_{K(\theta)} \widetilde{X}_\theta R_\theta \ . \tag{3.38}
$$

Some properties of the operator $\widetilde{\Pi}_\theta$ are collected in Section 3.5.2.

3.5.2 Computing $\left(DV_{K(\theta)} - \partial_\omega\right) M_\theta$

Using the expression (3.38) for $\partial_\omega \widetilde{Y}_\theta$, we obtain

$$\begin{aligned}
\left(DV_{K(\theta)} - \partial_\omega\right) M_\theta &= \left(DV_{K(\theta)} - \partial_\omega\right) \begin{bmatrix} 0 & 0 & 0 \\ \\ 0 & 0 & \widetilde{Y}_\theta \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & \left[\dfrac{\partial \underline{V}}{\partial \widetilde{x}}\right]_{K(\theta)} \widetilde{Y}_\theta \\ \\ 0 & 0 & \left[\dfrac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)} \widetilde{Y}_\theta - \partial_\omega \widetilde{Y}_\theta \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & \left[\dfrac{\partial \underline{V}}{\partial \widetilde{x}}\right]_{K(\theta)} \widetilde{Y}_\theta \\ \\ 0 & 0 & -2\widetilde{J}^{-1}_{K(\theta)} \widetilde{\Pi}_\theta \left[\dfrac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]^{\mathrm{sym}}_{K(\theta)} \widetilde{X}_\theta R_\theta \end{bmatrix} .
\end{aligned} \tag{3.39}$$

Below we collect several observations about the operators $\widetilde{\Pi}_\theta$ (3.37) and

$$\widetilde{J}^{-1}_{K(\theta)} \widetilde{\Pi}_\theta \widetilde{J}_{K(\theta)} = \mathbb{I}_{2n} - \widetilde{Y}_\theta \widetilde{X}_\theta^\top \widetilde{J}_{K(\theta)} \ ,$$

which follow easily from the definitions (3.20) and (3.21) of R_θ and $\widetilde{Y}_\theta$, and the properties (3.26) of the adapted basis:

- $\widetilde{\Pi}_\theta$ is symmetric,

$$\widetilde{\Pi}_\theta^\top = \left(\mathbb{I}_{2n} - \widetilde{X}_\theta R_\theta \widetilde{X}_\theta^\top\right)^\top = \widetilde{\Pi}_\theta \ ;$$

- both $\widetilde{\Pi}_\theta$ and $\widetilde{J}^{-1}_{K(\theta)}\widetilde{\Pi}_\theta\widetilde{J}_{K(\theta)}$ are idempotent:

$$\widetilde{\Pi}_\theta^2 = \widetilde{\Pi}_\theta \ , \qquad \left(\widetilde{J}^{-1}_{K(\theta)}\widetilde{\Pi}_\theta\widetilde{J}_{K(\theta)}\right)^2 = \widetilde{J}^{-1}_{K(\theta)}\widetilde{\Pi}_\theta\widetilde{J}_{K(\theta)} \ ;$$

- the n columns of the matrix $\widetilde{X}_\theta$ are in the kernel of $\widetilde{\Pi}_\theta$:

$$\widetilde{\Pi}_\theta\widetilde{X}_\theta = \widetilde{X}_\theta - \widetilde{X}_\theta R_\theta \widetilde{X}_\theta^\top \widetilde{X}_\theta = \widetilde{X}_\theta - \widetilde{X}_\theta = 0 \ ;$$

- the n columns of $\widetilde{X}_\theta$ are eigenvectors of $\widetilde{J}^{-1}_{K(\theta)}\widetilde{\Pi}_\theta\widetilde{J}_{K(\theta)}$ with eigenvalue 1, while the n columns of $\widetilde{Y}_\theta$ are in the kernel of $\widetilde{J}^{-1}_{K(\theta)}\widetilde{\Pi}_\theta\widetilde{J}_{K(\theta)}$:

$$\widetilde{J}^{-1}_{K(\theta)}\widetilde{\Pi}_\theta\widetilde{J}_{K(\theta)}\,\widetilde{X}_\theta = \widetilde{X}_\theta \ , \qquad \widetilde{J}^{-1}_{K(\theta)}\widetilde{\Pi}_\theta\widetilde{J}_{K(\theta)}\,\widetilde{Y}_\theta = 0 \ ; \tag{3.40}$$

- the properties (3.40) mean that $\widetilde{J}^{-1}_{K(\theta)}\widetilde{\Pi}_\theta\widetilde{J}_{K(\theta)}$ and $\left(\mathbb{I}_{2n} - \widetilde{J}^{-1}_{K(\theta)}\widetilde{\Pi}_\theta\widetilde{J}_{K(\theta)}\right)$ are projection operators corresponding to the splitting of the $2n$-dimensional space $T_{K(\theta)}\mathcal{Q}$ into two n-dimensional subspaces – one spanned by the columns of $\widetilde{X}_\theta$ (which is the intersection of $T_{K(\theta)}\mathcal{Q}$ and $\ker\Omega_{K(\theta)}$), and a second one spanned by the columns of $\widetilde{Y}_\theta$:

$$T_{K(\theta)}\mathcal{Q} \ = \ \mathrm{span}\left\{\big(\widetilde{X}_\theta\big)^{\bullet}{}_a\right\}_{a=1}^n \ \oplus \ \mathrm{span}\left\{\big(\widetilde{Y}_\theta\big)^{\bullet}{}_a\right\}_{a=1}^n \ .$$

3.5.3 Computing C_θ

Having computed $(DV_{K(\theta)} - \partial_\omega)M_\theta$, we will rewrite it in the form

$$(DV_{K(\theta)} - \partial_\omega)M_\theta = M_\theta C_\theta \ , \tag{3.41}$$

with

$$C_\theta = \begin{bmatrix} 0 & 0 & T_\theta \\ 0 & 0 & S_\theta \\ 0 & 0 & U_\theta \end{bmatrix} . \tag{3.42}$$

This representation of $(DV_{K(\theta)} - \partial_\omega)M_\theta$ plays a crucial role in Chapter 4. In the rest of this section we will compute the matrices T_θ, S_θ, and U_θ explicitly.

Comparing

$$\begin{aligned} M_\theta C_\theta &= \begin{bmatrix} \underline{Z}_\theta & \underline{X}_\theta & 0 \\ \widetilde{Z}_\theta & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} \begin{bmatrix} 0 & 0 & T_\theta \\ 0 & 0 & S_\theta \\ 0 & 0 & U_\theta \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & \underline{Z}_\theta T_\theta + \underline{X}_\theta S_\theta \\ 0 & 0 & \widetilde{Z}_\theta T_\theta + \widetilde{X}_\theta S_\theta + \widetilde{Y}_\theta U_\theta \end{bmatrix} \end{aligned}$$

with the expression (3.39) for $(DV_{K(\theta)} - \partial_\omega)M_\theta$, we see that the matrices T_θ, S_θ, and U_θ should satisfy the equations

$$\underline{Z}_\theta T_\theta + \underline{X}_\theta S_\theta = \left[\frac{\partial \underline{V}}{\partial \widetilde{x}}\right]_{K(\theta)} \widetilde{J}_{K(\theta)}^{-1} \widetilde{X}_\theta R_\theta \tag{3.43}$$

$$\widetilde{Z}_\theta T_\theta + \widetilde{X}_\theta S_\theta + \widetilde{Y}_\theta U_\theta = -2\widetilde{J}_{K(\theta)}^{-1} \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^{\mathrm{sym}} \widetilde{X}_\theta R_\theta . \tag{3.44}$$

Multiplying (3.44) separately by $\widetilde{Z}_\theta^\top \widetilde{J}_{K(\theta)}$, $\widetilde{X}_\theta^\top \widetilde{J}_{K(\theta)}$, and $\widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)}$ on the left and using (3.26) and the definition of R_θ (3.20), we obtain

$$\widetilde{Z}_\theta^\top \widetilde{X}_\theta R_\theta U_\theta = -2\widetilde{Z}_\theta^\top \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^{\mathrm{sym}} \widetilde{X}_\theta R_\theta \tag{3.45}$$

$$U_\theta = -2\widetilde{X}_\theta^\top \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^{\mathrm{sym}} \widetilde{X}_\theta R_\theta \tag{3.46}$$

$$-R_\theta \widetilde{X}_\theta^\top \widetilde{Z}_\theta T_\theta - S_\theta - R_\theta \widetilde{X}_\theta^\top \widetilde{J}_{K(\theta)}^{-1} \widetilde{X}_\theta R_\theta U_\theta = -2\widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^{\mathrm{sym}} \widetilde{X}_\theta R_\theta \ . \tag{3.47}$$

Recalling that $\widetilde{X}_\theta^\top \widetilde{\Pi}_\theta = 0$, we see from (3.46) that

$$U_\theta = 0 \ .$$

From (3.47) we express S_θ in terms of T_θ:

$$S_\theta = 2\widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^{\mathrm{sym}} \widetilde{X}_\theta R_\theta - R_\theta \widetilde{X}_\theta^\top \widetilde{Z}_\theta T_\theta \ .$$

Substitute this in (3.43):

$$\underline{Z}_\theta T_\theta + \underline{X}_\theta \left(2\widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^{\mathrm{sym}} \widetilde{X}_\theta R_\theta - R_\theta \widetilde{X}_\theta^\top \widetilde{Z}_\theta T_\theta\right) = \left[\frac{\partial \underline{V}}{\partial \widetilde{x}}\right]_{K(\theta)} \widetilde{J}_{K(\theta)}^{-1} \widetilde{X}_\theta R_\theta$$

to obtain

$$\left(\underline{Z}_\theta - \underline{X}_\theta R_\theta \widetilde{X}_\theta^\top \widetilde{Z}_\theta\right) T_\theta = \left(\left[\frac{\partial \underline{V}}{\partial \widetilde{x}}\right]_{K(\theta)} \widetilde{J}_{K(\theta)}^{-1} - 2\underline{X}_\theta \widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}}\right]_{K(\theta)}^{\mathrm{sym}}\right) \widetilde{X}_\theta R_\theta \ .$$

Finally,

$$T_\theta = \underline{\mathscr{Z}}_\theta^{-1} \left(\left[\frac{\partial \underline{V}}{\partial \widetilde{x}} \right]_{K(\theta)} \widetilde{Y}_\theta - 2\underline{X}_\theta \widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}} \right]_{K(\theta)}^{\mathrm{sym}} \widetilde{X}_\theta R_\theta \right) \ , \tag{3.48}$$

where we have set

$$\underline{\mathscr{Z}}_\theta := \underline{Z}_\theta - \underline{X}_\theta R_\theta \widetilde{X}_\theta^\top \widetilde{Z}_\theta = \underline{Z}_\theta + \underline{X}_\theta \widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{Z}_\theta \in \mathrm{M}_{d,d}(\mathbb{R}) \ . \tag{3.49}$$

The geometric meaning of $\underline{\mathscr{Z}}_\theta$ will become transparent after seeing the derivation of equation (3.53) below.

Substituting the expression (3.48) for T_θ into the expression for S_θ above, we obtain

$$\begin{aligned} S_\theta &= 2\widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}} \right]_{K(\theta)}^{\mathrm{sym}} \widetilde{X}_\theta R_\theta \\ &\quad - R_\theta \widetilde{X}_\theta^\top \widetilde{Z}_\theta \underline{\mathscr{Z}}_\theta^{-1} \left(\left[\frac{\partial \underline{V}}{\partial \widetilde{x}} \right]_{K(\theta)} \widetilde{J}_{K(\theta)}^{-1} - 2\underline{X}_\theta \widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}} \right]_{K(\theta)}^{\mathrm{sym}} \right) \widetilde{X}_\theta R_\theta \\ &= 2\widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}} \right]_{K(\theta)}^{\mathrm{sym}} \widetilde{X}_\theta R_\theta - R_\theta \widetilde{X}_\theta^\top \widetilde{Z}_\theta \underline{\mathscr{Z}}_\theta^{-1} \left[\frac{\partial \underline{V}}{\partial \widetilde{x}} \right]_{K(\theta)} \widetilde{J}_{K(\theta)}^{-1} \widetilde{X}_\theta R_\theta \\ &\quad + 2R_\theta \widetilde{X}_\theta^\top \widetilde{Z}_\theta \underline{\mathscr{Z}}_\theta^{-1} \underline{X}_\theta \widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}} \right]_{K(\theta)}^{\mathrm{sym}} \widetilde{X}_\theta R_\theta \ , \end{aligned}$$

which simplifies to

$$\begin{aligned} S_\theta &= -\widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{Z}_\theta \underline{\mathscr{Z}}_\theta^{-1} \left[\frac{\partial \underline{V}}{\partial \widetilde{x}} \right]_{K(\theta)} \widetilde{Y}_\theta \\ &\quad + 2\left(\mathbb{I}_n - \widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{Z}_\theta \underline{\mathscr{Z}}_\theta^{-1} \underline{X}_\theta \right) \widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \left[\frac{\partial \widetilde{V}}{\partial \widetilde{x}} \right]_{K(\theta)}^{\mathrm{sym}} \widetilde{X}_\theta R_\theta \ . \end{aligned} \tag{3.50}$$

We summarize our findings in the following

Lemma 3.6. *Let $(\mathcal{P}, \Omega)$ be an exact presymplectic manifold, $V \in \mathfrak{X}(\mathcal{P})$ be a presymplectic vector field, $K : \mathbb{T}^{d+n} \to \mathcal{P}$ be an invariant torus in the sense of Definition 3.1, and the matrix M_θ be defined by* (3.30)*.*

Then

$$\left(DV_{K(\theta)} - \partial_\omega\right) M_\theta = M_\theta \, C_\theta$$

with

$$C_\theta = \begin{bmatrix} 0 & 0 & T_\theta \\ 0 & 0 & S_\theta \\ 0 & 0 & 0 \end{bmatrix} ,$$

where T_θ and S_θ are given by (3.48) *and* (3.50)*, respectively.*

3.5.4 Factorizations of M_θ and M_θ^{-1}

In this section we use simple geometric ideas to derive a factorization of the change of basis matrix M_θ, which will also imply a factorization and an explicit expression for M_θ^{-1}. The derivation will elucidate the origin of the quantity $\underline{\mathscr{Z}}_\theta$ (3.49).

One can transform M_θ (3.30) by elementary operations to give it a simpler form. We perform this in a series of steps. Recall that

$$M_\theta = \begin{bmatrix} \underline{Z}_\theta & \underline{X}_\theta & 0 \\ \widetilde{Z}_\theta & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} .$$

- Since $K : \mathbb{T}^{d+n} \to \mathcal{P}$ is an embedding and because of our construction of the vectors $\widetilde{X}_\theta$ and $\widetilde{Y}_\theta$ in Section 3.3,

$$\operatorname{rank}\begin{bmatrix}\widetilde{X}_\theta & \widetilde{Y}_\theta\end{bmatrix} = 2n \ .$$

(In fact, the columns of the matrices $\widetilde{X}_\theta$ and $\widetilde{Y}_\theta$ form a symplectic basis of the $2n$-dimensional symplectic subspace $T_{K(\theta)}\mathcal{Q} \subseteq T_{K(\theta)}\mathcal{P}$ which is a realization of the factorspace $T_{K(\theta)}\mathcal{P}/\ker\Omega_{K(\theta)}$, as explained in Section 2.1.2.) This implies that the lower part of the matrix M_θ, namely, the matrix

$$\begin{bmatrix}\widetilde{Z}_\theta & \widetilde{X}_\theta & \widetilde{Y}_\theta\end{bmatrix} \in \mathrm{M}_{2n,d+2n}(\mathbb{R}) \ ,$$

is of maximum rank. Hence, the d columns of the matrix $\widetilde{Z}_\theta \in \mathrm{M}_{2n,d}(\mathbb{R})$ are linear combinations of the columns of $\widetilde{X}_\theta$ and $\widetilde{Y}_\theta$. This fact can be written in matrix notations as

$$\widetilde{Z}_\theta = \widetilde{X}_\theta A'_\theta + \widetilde{Y} B'_\theta \ , \qquad A'_\theta, B'_\theta \in \mathrm{M}_{n,d}(\mathbb{R}) \ ; \tag{3.51}$$

the matrices A'_θ and B'_θ are introduced temporarily and will be used only in this section. Multiply (3.51) on the left by $\widetilde{X}_\theta^\top \widetilde{J}_{K(\theta)}$ and use the properties (3.26) of the adapted basis to get $B'_\theta = 0$, so that

$$\widetilde{Z}_\theta = \widetilde{X}_\theta A'_\theta \ .$$

On the other hand, multiplying (3.51) on the left by $\widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)}$, we obtain

$$\widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{Z}_\theta = \widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{X}_\theta A'_\theta = -\mathbb{I}_n A'_\theta = -A'_\theta \ .$$

Therefore M_θ can be written as

$$M_\theta = \begin{bmatrix} \underline{Z}_\theta & \underline{X}_\theta & 0 \\ \widetilde{X}_\theta A'_\theta & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} , \qquad A'_\theta = -\widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{Z}_\theta = R_\theta \widetilde{X}_\theta^\top \widetilde{Z}_\theta \ . \quad (3.52)$$

- Add the second group of n columns of M_θ (i.e., the matrix X_θ) multiplied by $-A'_\theta$ to the first group of d columns (i.e., Z_θ) to eliminate the "symplectic component" of the vector Z_θ:

$$M_\theta = \begin{bmatrix} \underline{Z}_\theta & \underline{X}_\theta & 0 \\ \widetilde{X}_\theta A'_\theta & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} \sim \begin{bmatrix} \underline{Z}_\theta - \underline{X}_\theta A'_\theta & \underline{X}_\theta & 0 \\ 0 & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} .$$

The $d \times d$ block in the upper left corner of the matrix in the right-hand side is exactly the expression

$$\underline{\mathscr{Z}}_\theta = \underline{Z}_\theta - \underline{X}_\theta A'_\theta = \underline{Z}_\theta - \underline{X}_\theta R_\theta \widetilde{X}_\theta^\top \widetilde{Z}_\theta \in \mathrm{M}_{d,d}(\mathbb{R}) \ ,$$

introduced in (3.49). The operation above is equivalent to a matrix multiplication:

$$\begin{bmatrix} \underline{\mathscr{Z}}_\theta & \underline{X}_\theta & 0 \\ 0 & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} = \begin{bmatrix} \underline{Z}_\theta & \underline{X}_\theta & 0 \\ \widetilde{X}_\theta A'_\theta & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} \begin{bmatrix} \mathbb{I}_d & 0 & 0 \\ -A'_\theta & \mathbb{I}_n & 0 \\ 0 & 0 & \mathbb{I}_n \end{bmatrix} . \qquad (3.53)$$

- Since the matrix in the left-hand side of (3.53) has full rank, its upper left corner, $\underline{\mathscr{Z}}_\theta \in \mathrm{M}_{d,d}(\mathbb{R})$, has full rank, so that it is invertible. Using this block to eliminate the block $\underline{X}_\theta \in \mathrm{M}_{d,n}(\mathbb{R})$ (i.e., the first d rows of the matrix X_θ), we obtain

$$M_\theta \sim \begin{bmatrix} \underline{\mathscr{Z}}_\theta & \underline{X}_\theta & 0 \\ 0 & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} \sim \begin{bmatrix} \underline{\mathscr{Z}}_\theta & 0 & 0 \\ 0 & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix}.$$

In terms of matrix multiplication this can be written as

$$\begin{bmatrix} \underline{\mathscr{Z}}_\theta & 0 & 0 \\ 0 & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} = \begin{bmatrix} \underline{\mathscr{Z}}_\theta & \underline{X}_\theta & 0 \\ 0 & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} \begin{bmatrix} \mathbb{I}_d & -\underline{\mathscr{Z}}_\theta^{-1}\underline{X}_\theta & 0 \\ 0 & \mathbb{I}_n & 0 \\ 0 & 0 & \mathbb{I}_n \end{bmatrix}.$$

- Finally, we have

$$M_\theta \sim \begin{bmatrix} \underline{\mathscr{Z}}_\theta & 0 & 0 \\ 0 & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} = \begin{bmatrix} \mathbb{I}_d & 0 & 0 \\ 0 & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} \begin{bmatrix} \underline{\mathscr{Z}}_\theta & 0 & 0 \\ 0 & \mathbb{I}_n & 0 \\ 0 & 0 & \mathbb{I}_n \end{bmatrix},$$

  which implies that

$$\det M_\theta = \det\left(\underline{\mathscr{Z}}_\theta\right) \det\left[\widetilde{X}_\theta \ \widetilde{Y}_\theta\right].$$

The inverse matrix of $\left[\widetilde{X}_\theta \ \widetilde{Y}_\theta\right]$ is

$$\left[\widetilde{X}_\theta \ \widetilde{Y}_\theta\right]^{-1} = \begin{bmatrix} -\widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \widetilde{J}_{K(\theta)} \\ \widetilde{X}_\theta^\top \widetilde{J}_{K(\theta)} \end{bmatrix} = \begin{bmatrix} -\widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \\ \widetilde{X}_\theta^\top \end{bmatrix} \widetilde{J}_{K(\theta)}.$$

To summarize, we have found the factorization

$$\begin{bmatrix} \underline{\mathscr{Z}}_\theta & 0 & 0 \\ 0 & \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix} = M_\theta \begin{bmatrix} \mathbb{I}_d & 0 & 0 \\ -A'_\theta & \mathbb{I}_n & 0 \\ 0 & 0 & \mathbb{I}_n \end{bmatrix} \begin{bmatrix} \mathbb{I}_d & -\underline{\mathscr{Z}}_\theta^{-1}\underline{X}_\theta & 0 \\ 0 & \mathbb{I}_n & 0 \\ 0 & 0 & \mathbb{I}_n \end{bmatrix},$$

from which we can obtain the inverse matrix of M_θ:

$$\begin{aligned} M_\theta^{-1} &= \begin{bmatrix} \mathbb{I}_d & 0 & 0 \\ -A'_\theta & \mathbb{I}_n & 0 \\ 0 & 0 & \mathbb{I}_n \end{bmatrix} \begin{bmatrix} \mathbb{I}_d & -\underline{\mathscr{Z}}_\theta^{-1}\underline{X}_\theta & 0 \\ 0 & \mathbb{I}_n & 0 \\ 0 & 0 & \mathbb{I}_n \end{bmatrix} \begin{bmatrix} \underline{\mathscr{Z}}_\theta^{-1} & 0 \\ 0 & \begin{bmatrix} \widetilde{X}_\theta & \widetilde{Y}_\theta \end{bmatrix}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbb{I}_d & 0 & 0 \\ -A'_\theta & \mathbb{I}_n & 0 \\ 0 & 0 & \mathbb{I}_n \end{bmatrix} \begin{bmatrix} \mathbb{I}_d & -\underline{\mathscr{Z}}_\theta^{-1}\underline{X} & 0 \\ 0 & \mathbb{I}_n & 0 \\ 0 & 0 & \mathbb{I}_n \end{bmatrix} \begin{bmatrix} \underline{\mathscr{Z}}_\theta^{-1} & 0 \\ 0 & -\widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \widetilde{J}_{K(\theta)} \\ 0 & \widetilde{X}_\theta^\top \widetilde{J}_{K(\theta)} \end{bmatrix} \qquad (3.54) \\ &= \begin{bmatrix} \underline{\mathscr{Z}}_\theta^{-1} & \underline{\mathscr{Z}}_\theta^{-1}\underline{X}_\theta \widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \widetilde{J}_{K(\theta)} \\ -A'_\theta \underline{\mathscr{Z}}_\theta^{-1} & -\left(\mathbb{I}_n + A'_\theta \underline{\mathscr{Z}}_\theta^{-1}\underline{X}_\theta\right) \widetilde{Y}_\theta^\top \widetilde{\Pi}_\theta \widetilde{J}_{K(\theta)} \\ 0 & \widetilde{X}_\theta^\top \widetilde{J}_{K(\theta)} \end{bmatrix}, \end{aligned}$$

where $\widetilde{\Pi}_\theta$, $\underline{\mathscr{Z}}_\theta$, and A'_θ were introduced in (3.37), (3.49), and (3.52), respectively.

3.5.5 Other factorizations of M_θ and M_θ^{-1}

Later we will use another representation of M_θ^{-1} from the Lemma below.

Lemma 3.7. *If the $(d+2n)\times(d+2n)$ matrices Q_θ and W_θ are defined by*

$$Q_\theta := \begin{bmatrix} \mathbb{I}_d & 0 \\ 0 & \widetilde{X}_{0,\theta}^\top \\ 0 & \widetilde{Y}_{0,\theta}^\top \end{bmatrix} \begin{bmatrix} \mathbb{I}_d & 0 \\ 0 & \widetilde{J}_{K_0(\theta)} \end{bmatrix} = \begin{bmatrix} \mathbb{I}_d & 0 \\ 0 & \widetilde{X}_\theta^\top \widetilde{J}_{K(\theta)} \\ 0 & \widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)} \end{bmatrix} , \tag{3.55}$$

$$W_\theta = \begin{bmatrix} \underline{Z}_\theta & \underline{X}_\theta & 0 \\ 0 & 0 & \mathbb{I}_n \\ \widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{Z}_\theta & -\mathbb{I}_n & \widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)} \widetilde{Y}_\theta \end{bmatrix} , \tag{3.56}$$

then the following identity holds:

$$Q_\theta M_\theta = W_\theta \ . \tag{3.57}$$

This, in particular, implies that the matrix M_θ (3.30) *is invertible if and only if the matrix W_θ is invertible.*

Proof. The columns of $\widetilde{X}_\theta$ and $\widetilde{Y}_\theta$ form a (symplectic) basis of $\mathbb{R}^{2n}$, which implies that the rows of $\widetilde{X}_\theta^\top$ and $\widetilde{Y}_\theta^\top$ form a basis of $\mathbb{R}^{2n}$. Since $\widetilde{J}_{K(\theta)}$ is an invertible matrix (it corresponds to the symplectic form $\widetilde{\Omega}$ on $\mathcal{Q}$, recall (2.6) and (2.8)), the rows of $\widetilde{X}_\theta^\top \widetilde{J}_{K(\theta)}$ and $\widetilde{Y}_\theta^\top \widetilde{J}_{K(\theta)}$ from a basis of $\mathbb{R}^{2n}$, so that the matrix Q_θ given by (3.55) is invertible. The identity (3.57) follows directly from (3.26). $\square$

An immediate consequence of Lemma 3.7 is the following factorization of M_θ^{-1}:

$$M_\theta^{-1} = W_\theta^{-1} Q_\theta \ . \tag{3.58}$$

Chapter 4

Approximate Solutions

In this section we will examine what happens when K is merely an approximate solution as defined below. We will build off of the results in Chapter 3 for true solutions to show that similar results still hold for approximate solutions. We start with the definition for approximate solution.

Definition 4.1. *Let $\mathcal{P}$ be an exact presymplectic manifold, $V_\lambda \in \mathfrak{X}(\mathcal{P})$ be a $(d+2n)$-parameter family of presymplectic vector fields, $\omega \in \mathcal{D}(\gamma, \sigma)$ be a Diophantine vector of dimension $(d+n)$, and*

$$K_0 \colon \mathbb{T}^{d+n} \to \mathcal{P}$$

be an embedding. For a value λ_0 of the parameter λ, define the error,

$$e_{0,\theta} := V_{\lambda_0, K_0(\theta)} - \partial_\omega K_{0,\theta} \in T_{K_0(\theta)}\mathcal{P} \cong \mathbb{R}^{d+2n} \ . \tag{4.1}$$

If some appropriately defined norm of e_0 is sufficiently small, then we say that K_0 is an ***approximate solution****.*

We will usually consider e_0 as a map

$$e_0 : \mathbb{T}^{d+n} \to \mathbb{R}^{d+2n} \ ,$$

whose derivative,

$$De_0 : \mathbb{T}^{d+n} \to \mathrm{M}_{d+2n,d+n}(\mathbb{R}) \ , \tag{4.2}$$

is given by

$$De_{0,\theta} = \big(DV_{\lambda_0,\, K_0(\theta)} - \partial_\omega\big)\, DK_{0,\theta} \in \mathrm{M}_{d+2n,d+n}(\mathbb{R}) \ . \tag{4.3}$$

Note that in (4.3), $DV_{\lambda_0,\, K_0(\theta)}$ stands for the derivative of the vector field V_{λ_0} *with respect to the spatial variables* $x \in \mathbb{R}^{d+2n}$:

$$DV_{\lambda_0,K_0(\theta)} = \Big[\big(DV_{\lambda_0,K_0(\theta)}\big)^A{}_B\Big] = \left[\left.\frac{\partial V^A_{\lambda_0,x}}{\partial x^B}\right|_{x=K_0(\theta)}\right] \in \mathrm{M}_{d+2n,d+2n}(\mathbb{R}) \ .$$

Recall that the presymplecticity of the family V_λ imply that the matrix $DV_{\lambda_0,K_0(\theta)}$ is block upper triangular (3.29), and its component $\left[\dfrac{\partial \widetilde{V}_{\lambda_0}}{\partial \widetilde{x}}\right]_{K_0(\theta)}$ satisfies (3.28).

4.1 Approximately isotropic tori

In Lemma 3.5 we showed that if K_θ is a true solution (i.e., if (3.1) is satisfied), then the invariant manifold $\mathcal{K}$ (2.18) is isotropic, i.e., $K^*\Omega = 0$. The analogous result for this chapter will be that if $K_{0\,,\,\theta}$ is an approximate solution, then K_0 is **approximately isotropic**, i.e., $K_0^*\Omega$ is small.

Lemma 4.2. *Let $\mathcal{P}$ be an exact presymplectic manifold, $V_\lambda \in \mathfrak{X}(\mathcal{P})$ be a $(d+2n)$-parameter family of presymplectic analytic vector fields, and $K_0 \in \mathcal{W}_\rho$ (2.13) be* *an approximate solution with Diophantine frequency* $\omega \in \mathcal{D}(\gamma, \sigma)$*. Assume that* V_λ *extends holomorphically to some complex neighborhood* $\mathcal{B}_r$ (2.27) *of the image of* $\mathbb{T}^{d+n}_\rho$ *under* K_0*, for some* $r > 0$*. Let*

$$L_{0,\theta} : T_\theta \mathbb{T}^{d+n} \to \mathbb{T}_\theta \mathbb{T}^{d+n}$$

be the matrix representation of the pull-back $(K_0^* \Omega)_\theta$ *as in* (3.3) *and* (3.4)*:*

$$L_{0,\theta} = DK_{0,\theta}^\top \, J_{K_0(\theta)} \, DK_{0,\theta} \ . \tag{4.4}$$

Then there exists a constant $C > 0$ *depending on* d*,* n*,* σ*,* ρ*,* $\|DK_0\|_\rho$*,* $|V_{\lambda_0}|_{C^1, \mathcal{B}_r}$*, and* $|J|_{C^1, \mathcal{B}_r}$*, such that for every* δ *satisfying*

$$0 < \delta < \frac{\rho}{2} \ ,$$

the following bound holds:

$$\|L_0\|_{\rho - 2\delta} < C \gamma^{-1} \delta^{-(\sigma+1)} \, \|e_0\|_\rho \ . \tag{4.5}$$

Proof. Consider the directional derivative of the linear operator $L_{0,\theta}$ (4.4). Using

(4.3), we obtain

$$
\begin{aligned}
\partial_\omega L_{0,\theta} &= \partial_\omega\big(DK_{0,\theta}^\top\, J_{K_0(\theta)}\, DK_{0,\theta}\big) \\
&= \partial_\omega\big(DK_{0,\theta}^\top\big) J_{K_0(\theta)}\, DK_{0,\theta} + DK_{0,\theta}^\top\, \partial_\omega\big(J_{K_0(\theta)}\big)\, DK_{0,\theta} + DK_{0,\theta}^\top\, J_{K_0(\theta)}\, \partial_\omega\big(DK_{0,\theta}\big) \\
&= \big(DV_{\lambda_0,\, K_0(\theta)}\, DK_{0,\theta} - De_{0,\theta}\big)^\top J_{K_0(\theta)}\, DK_{0,\theta} \\
&\quad + DK_{0,\theta}^\top\, DJ_{K_0(\theta)}\, \left(V_{\lambda_0,K_0(\theta)} - e_{0,\theta}\right)\, DK_{0,\theta} \\
&\quad + DK_{0,\theta}^\top\, J_{K_0(\theta)}\, \big(DV_{\lambda_0,\, K_0(\theta)}\, DK_{0,\theta} - De_{0,\theta}\big) \\
&= DK_{0,\theta}^\top\, \big(DV_{\lambda_0,\, K_0(\theta)}^\top\, J_{K_0(\theta)} + DJ_{K_0(\theta)}\, V_{\lambda_0,\, K_0(\theta)} + J_{K_0(\theta)}\, DV_{\lambda_0,\, K_0(\theta)}\big)\, DK_{0,\theta} \\
&\quad - \left(De_{0,\theta}^\top\, J_{K_0(\theta)}\, DK_{0,\theta} + DK_{0,\theta}^\top\, DJ_{K_0(\theta)}\, e_{0,\theta}\, DK_{0,\theta} + DK_{0,\theta}^\top\, J_{K_0(\theta)}\, De_{0,\theta}\right) \\
&= -\left(De_{0,\theta}^\top\, J_{K_0(\theta)}\, DK_{0,\theta} + DK_{0,\theta}^\top\, DJ_{K_0(\theta)}\, e_{0,\theta}\, DK_{0,\theta} + DK_{0,\theta}^\top\, J_{K_0(\theta)}\, De_{0,\theta}\right)\ .
\end{aligned}
$$

In the last step we used the identity (2.26) coming from the presymplecticity of V_{λ_0}. From this and the Cauchy bounds (2.15) we obtain

$$
\|\partial_\omega L_0\|_{\rho-\delta} \leq C_1\|e_0\|_{\rho-\delta} + C_2\|De_0\|_{\rho-\delta} \leq C\delta^{-1}\|e_0\|_\rho\ . \tag{4.6}
$$

Although K_0 is only an approximate solution, the exactness of the presymplectic form Ω implies that the average of L_0 over $\mathbb{T}^{d+n}$ vanishes exactly:

$$
\operatorname{avg}(L_0) = 0\ . \tag{4.7}
$$

The proof of this repeats the part of the proof of Lemma 3.5 between equations (3.4) and (3.5), with K replaced by K_0. Because of (4.7), we can apply the

Rüssmann estimate (2.16) to obtain

$$\|L_0\|_{\rho-2\delta} \leq C\gamma^{-1}\delta^{-\sigma} \|\partial_\omega L_0\|_{\rho-\delta} \leq C\gamma^{-1}\delta^{-(\sigma+1)}\|e_0\|_\rho ,$$

where in the last step we used (4.6).

□

4.2 Derivation of the linearized equation

Given a family of presymplectic vector fields V_λ, the implicit equation

$$V_{\lambda,K_\theta} = \partial_\omega K_\theta$$

can be difficult to solve for an embedding $K\colon \mathbb{T}^{d+n} \to \mathcal{P}$ and a value $\bar{\lambda}$ of the parameter that satisfies the equation for $\lambda = \bar{\lambda}$. So instead of solving this equation directly for K and λ, we will start with an approximate solution and construct an iterative process that will produce better approximate solutions that converge to a true solution. As a result of this iterative process we will find a sequence of pairs (λ_j, K_j) that will converge to a pair $(\lambda_\infty, K_\infty)$ such that

$$V_{\lambda_\infty,K_\infty} = \partial_\omega K_\infty .$$

Let (λ_0, K_0) be the initial approximate pair. Define the error e_0 as in (4.1), and assume that its norm is small. Since K_0 is not a true solution, we will be interested in constructing an improved approximate solution by adding correction terms to both λ_0 and K_0. Define the vector $\varepsilon_0 \in \mathbb{R}^{d+2n}$ and the function

$$\Delta_0 \colon \mathbb{T}^{d+n} \to \mathcal{P}$$

to be the correction terms for λ_0 and K_0, respectively, so that the pair

$$(\lambda_1, K_1) := (\lambda_0 + \varepsilon_0, K_0 + \Delta_0)$$

is a better approximate solution, in the sense that the norm of the error

$$e_{1,\theta} := V_{\lambda_1, K_1(\theta)} - \partial_\omega K_{1,\theta}$$

will be smaller than $\|e_0\|$.

In general, define

$$e_{j,\theta} := V_{\lambda_j, K_j(\theta)} - \partial_\omega K_{j,\theta} \ , \tag{4.8}$$

and let ε_j and Δ_j be $(j+1)$st correction terms, i.e.,

$$\lambda_{j+1} := \lambda_j + \varepsilon_j \ , \qquad K_{j+1,\theta} := K_{j,\theta} + \Delta_{j,\theta} \ . \tag{4.9}$$

In the iterative process we will use the Cauchy estimate (2.15) and the Rüssmann estimate (2.16), so it is clear that the domain of the embedding K_j will shrink as j increases. This phenomenon, called **loss of domain**, leads us towards a precarious situation. Could the domain run out before the process converges? The key idea here is that in the Newton method for solving nonlinear equations, the errors decay **quadratically**, i.e.,

$$\|e_{j+1}\| < C \|e_j\|^2$$

for some constant C (at the moment we ignore the domains over which the norms are taken). The quadratic convergence of the errors and a careful selection of how much domainto give up at each step is enough to ensure that our method converges before the domain runs out.

In the rest of this section we will derive a linear equation for the corrections ε_j and Δ_j. Define the operator $\mathcal{F}$ acting on a pair (λ, K) by

$$\mathcal{F}[\lambda, K](\theta) := V_{\lambda, K(\theta)} - \partial_\omega K_\theta \ .$$

Then a true solution (λ, K) would satisfy $\mathcal{F}[\lambda, K](\theta) = 0$. With this notation,

$$e_{j,\theta} = \mathcal{F}[\lambda_j, K_j](\theta) \ .$$

Therefore

$$\begin{aligned}
e_{j+1,\theta} &= \mathcal{F}[\lambda_{j+1}, K_{j+1}](\theta) = \mathcal{F}[\lambda_j + \varepsilon_j, K_j + \Delta_j](\theta) \\
&= V_{\lambda_j+\varepsilon_j,\, K_j(\theta)+\Delta_j(\theta)} - \partial_\omega\big(K_{j,\theta} + \Delta_{j,\theta}\big) \\
&= V_{\lambda_j,\, K_j(\theta)} + DV_{\lambda_j,\, K_j(\theta)}\, \Delta_{j,\theta} + \left[\frac{\partial V_\lambda}{\partial \lambda}\right]_{\lambda_j, K_j(\theta)} \varepsilon_j \\
&\quad - \partial_\omega K_{j,\theta} - \partial_\omega \Delta_{j,\theta} + \mathcal{O}(|\Delta_j, \varepsilon_j|^2) \\
&= \mathcal{F}[\lambda_j, K_j](\theta) + DV_{\lambda_j,\, K_j(\theta)}\, \Delta_{j,\theta} - \partial_\omega \Delta_{j,\theta} + \left[\frac{\partial V_\lambda}{\partial \lambda}\right]_{\lambda_j, K_j(\theta)} \varepsilon_j + \mathcal{O}(|\Delta_j, \varepsilon_j|^2) \\
&= e_{j,\theta} + DV_{\lambda_j,\, K_j(\theta)}\, \Delta_{j,\theta} - \partial_\omega \Delta_{j,\theta} + \left[\frac{\partial V_\lambda}{\partial \lambda}\right]_{\lambda_j, K_j(\theta)} \varepsilon_j + \mathcal{O}(|\Delta_j, \varepsilon_j|^2) \ .
\end{aligned} \tag{4.10}$$

So, if we can find ε_j and Δ_j such that

$$\left(DV_{\lambda_j,\,K_j(\theta)} - \partial_\omega\right)\Delta_{j,\theta} = -e_{j,\theta} - \left[\frac{\partial V_\lambda}{\partial \lambda}\right]_{\lambda_j, K_j(\theta)} \varepsilon_j \ , \tag{4.11}$$

then all of the terms that depend linearly on λ_j and Δ_j in the right-hand side of (4.10) will cancel out, so that only terms that are quadratic and higher powers in λ_j and Δ_j will remain, which will ensure that the scheme is quadratically convergent.

The system (4.11) of $(d + 2n)$ equations for the unknown corrections ε_j and Δ_j to the parameter λ_j and the embedding K_j is a linear algebraic equation with respect to the components of the vector ε_j, and a linear first-order partial differential equation with respect to the unknown functions Δ_j. Since the matrix $DV_{\lambda_j,\,K_j(\theta)} \in \mathrm{M}_{d+2n,d+2n}(\mathbb{R})$ is of a general form, is not easy to solve the system (4.11) and to obtain estimates on the size of its solution. A powerful idea that will help us solve (4.11) is to use the underlying geometry and dynamics. We will use the basis introduced in Section 3.3, with the help of the change of basis matrix introduced in Section 3.4; the calculations in Section 3.5 will be very useful.

We undertake this strategy for solving (4.11) in Section 4.3.

4.3 Solving the linearized equation

4.3.1 Geometric considerations

Now we employ a geometric strategy for solving the equation (4.11) for the unknown constants $\varepsilon_j \in \mathbb{R}^{d+2n}$ and functions $\Delta_j : \mathbb{T}^{d+n} \to \mathbb{R}^{d+2n}$. Instead of using a general subscript $j \in \{0, 1, 2, \ldots\}$, we will write a subscript 0 to denote the approximate solution, and with replace 0 with j after the end of the derivation.

We rewrite (4.11) in the form

$$\left(DV_{\lambda_0,\,K_0(\theta)} - \partial_\omega\right)\Delta_{0,\theta} = -e_{0,\theta} - \left[\frac{\partial V_\lambda}{\partial \lambda}\right]_{\lambda_0, K_0(\theta)} \varepsilon_0\ . \tag{4.12}$$

To utilize the geometry behind equation (4.12), we introduce an adapted basis in $\mathbb{R}^{d+2n}$, so that instead of the unknown function Δ_0 we introduce the unknown function

$$\xi_0 : \mathbb{T}^{d+n} \to \mathbb{R}^{d+2n}$$

through the linear change of basis

$$\Delta_{0,\theta} =: M_{0,\theta}\,\xi_{0,\theta}\ . \tag{4.13}$$

The change of basis matrix $M_{0,\theta} \in \mathrm{M}_{d+2n,d+2n}(\mathbb{R})$ is constructed similarly to the matrix M_θ in (3.30), but by using the approximate value λ_0 and the approximate embedding K_0. Namely, given an approximate invariant torus K_0 (4.12), which we treat as a map $K_0\colon \mathbb{T}^{d+n} \to \mathbb{R}^{d+2n}$, we define

$$\begin{bmatrix} \underline{Z}_{0,\theta} & \underline{X}_{0,\theta} \\ \widetilde{Z}_{0,\theta} & \widetilde{X}_{0,\theta} \end{bmatrix} := [Z_{0,\theta}\ X_{0,\theta}] := DK_{0,\theta} \in \mathrm{M}_{d+2n,d+n}(\mathbb{R})$$

as in (3.18) and (3.19), the inverse Gramian of the columns of X_0,

$$R_{0,\theta} := \left(\widetilde{X}_{0,\theta}^\top \widetilde{X}_{0,\theta}\right)^{-1} \in \mathrm{M}_{n,n}(\mathbb{R}) \tag{4.14}$$

as in (3.20), the matrices $\widetilde{Y}_{0,\theta}$ and $Y_{0,\theta}$,

$$\widetilde{Y}_{0,\theta} := \widetilde{J}^{-1}_{K_0(\theta)} \widetilde{X}_{0,\theta} R_{0,\theta} \in \mathrm{M}_{2n,n}(\mathbb{R}) \ , \qquad Y_{0,\theta} := \begin{bmatrix} 0 \\ \widetilde{Y}_{0,\theta} \end{bmatrix} \in \mathrm{M}_{d+2n,n}(\mathbb{R}) \tag{4.15}$$

as in (3.21) and (3.22), and the approximate change of basis matrix

$$M_{0,\theta} := [DK_{0,\theta} \ \ Y_{0,\theta}] = [Z_{0,\theta} \ X_{0,\theta} \ Y_{0,\theta}] = \begin{bmatrix} \underline{Z}_{0,\theta} & \underline{X}_{0,\theta} & 0 \\ \widetilde{Z}_{0,\theta} & \widetilde{X}_{0,\theta} & \widetilde{Y}_{0,\theta} \end{bmatrix} \in \mathrm{M}_{d+2n,d+2n}(\mathbb{R}) \tag{4.16}$$

as in (3.30).

As before, we think of the column vectors $(Z_{0,\theta})^{\bullet}{}_{\mu}$, $(X_{0,\theta})^{\bullet}{}_{a}$, and $(Y_{0,\theta})^{\bullet}{}_{a}$ as vectors in $T_{K_0(\theta)}\mathcal{P}$. If the map K_0 is close to the true solution K, then these vectors still form a basis of $T_{K_0(\theta)}\mathcal{P}$ as in the true case (recall (3.25)):

$$\mathrm{span}\left\{ \left\{ (Z_{0,\theta})^{\bullet}{}_{\mu} \right\}_{\mu=1}^{d}, \left\{ (X_{0,\theta})^{\bullet}{}_{a} \right\}_{a=1}^{n}, \left\{ (Y_{0,\theta})^{\bullet}{}_{a} \right\}_{a=1}^{n} \right\} = T_{K_0(\theta)}\mathcal{P} \cong \mathbb{R}^{d+2n} \ . \tag{4.17}$$

By construction, it is also clear that the columns of $Z_{0,\theta}$ and $X_{0,\theta}$ span the tangent space to the approximately invariant torus $\mathcal{K}_0 := K_0(\mathbb{T}^{d+n})$:

$$\mathrm{span}\left\{ \left\{ (Z_{0,\theta})^{\bullet}{}_{\mu} \right\}_{\mu=1}^{d}, \left\{ (X_{0,\theta})^{\bullet}{}_{a} \right\}_{a=1}^{n} \right\} = T_{K_0(\theta)}\mathcal{K}_0 \ . \tag{4.18}$$

Unlike the case of a true solution, however, the manifold $\mathcal{K}_0$ (and, therefore its tangent bundle) is not invariant with respect to the flow of the presymplectic vector field V_{λ_0}. Another fact to notice is that the kernel of the presymplectic form at $K_0(\theta)$ is generally not a subspace of the tangent space $T_{K(\theta)}\mathcal{K}_0$ to the manifold $\mathcal{K}_0$ at the point $K_0(\theta)$.

We make the substitution (4.13) in the variational equation (4.12) and obtain the following equation for the new unknown function $\xi_{0,\theta}$:

$$\left(DV_{\lambda_0,K_0(\theta)}M_{0,\theta} - \partial_\omega M_{0,\theta}\right)\xi_{0,\theta} - M_{0,\theta}\partial_\omega \xi_{0,\theta} = -e_{0,\theta} - \left[\frac{\partial V_\lambda}{\partial \lambda}\right]_{\lambda_0,K_0(\theta)} \varepsilon_0 \ .$$

Assuming that the matrix $M_{0,\theta}$ is invertible, we rewrite this as

$$M_{0,\theta}^{-1}\left(DV_{\lambda_0,K_0(\theta)}M_{0,\theta} - \partial_\omega M_{0,\theta}\right)\xi_{0,\theta} - \partial_\omega \xi_{0,\theta} = -M_{0,\theta}^{-1}\left(e_{0,\theta} + \left[\frac{\partial V_\lambda}{\partial \lambda}\right]_{\lambda_0,K_0(\theta)} \varepsilon_0\right) \tag{4.19}$$

Our immediate goal is to transform the coefficient of $\xi_{0,\theta}$ in this equation to a simpler form.

4.3.2 "Big" and "small" parts of the coefficients

Directly from the definition (4.16) of $M_{0,\theta}$, we obtain

$$\left(DV_{\lambda_0,K_0(\theta)} - \partial_\omega\right)M_{0,\theta} = \left[De_{0,\theta} \quad \left(DV_{\lambda_0,K_0(\theta)} - \partial_\omega\right)Y_{0,\theta}\right] \ .$$

To compute $\partial_\omega Y_{0,\theta}$, one can easily modify the computations in the derivation of the expression (3.38) for $\partial_\omega Y_\theta$ in the true solution case (Section 3.5). Here are some intermediate results: equations (3.33), (3.34), and (3.36) are replaced by

$$
\begin{aligned}
\partial_\omega\big(\widetilde{J}^{-1}_{K_0(\theta)}\big) &= \left[\frac{\partial \widetilde{V}_{\lambda_0}}{\partial \widetilde{x}}\right]_{K_0(\theta)} \widetilde{J}^{-1}_{K_0(\theta)} + \widetilde{J}^{-1}_{K_0(\theta)} \left[\frac{\partial \widetilde{V}_{\lambda_0}}{\partial \widetilde{x}}\right]^{\top}_{K_0(\theta)} + \widetilde{J}^{-1}_{K_0(\theta)}\, D\widetilde{J}_{K_0(\theta)}\, e_{0,\theta}\, \widetilde{J}^{-1}_{K_0(\theta)} \ , \\
\partial_\omega \widetilde{X}_{0,\theta} &= \left[\frac{\partial \widetilde{V}_{\lambda_0}}{\partial \widetilde{x}}\right]_{K_0(\theta)} \widetilde{X}_{0,\theta} - \left[\frac{\partial \widetilde{e}_{0,\theta}}{\partial \boldsymbol{\theta}}\right] \ , \\
\partial_\omega R_{0,\theta} &= -2R_{0,\theta} \widetilde{X}^{\top}_{0,\theta} \left[\frac{\partial \widetilde{V}_{\lambda_0}}{\partial \widetilde{x}}\right]^{\mathrm{sym}}_{K_0(\theta)} \widetilde{X}_{0,\theta} R_{0,\theta} \\
&\qquad + R_{0,\theta} \left\{ \left[\frac{\partial \widetilde{e}_{0,\theta}}{\partial \boldsymbol{\theta}}\right]^{\top} \widetilde{X}_{0,\theta} + \widetilde{X}^{\top}_{0,\theta} \left[\frac{\partial \widetilde{e}_{0,\theta}}{\partial \boldsymbol{\theta}}\right] \right\} R_{0,\theta} \ .
\end{aligned}
$$

In the right-hand side of the first equation, $D\widetilde{J}_{K_0(\theta)}\, e_{0,\theta}$ stands for the $2n \times 2n$ matrix with entries

$$
\Big(D\widetilde{J}_{K_0(\theta)}\, e_{0,\theta}\Big)^{i}{}_{j} := \sum_{A=1}^{d+2n} \left.\frac{\partial \widetilde{J}^{i}{}_{j}}{\partial x^A}\right|_{K_0(\theta)} e^{A}_{0,\theta}
$$

(compare with (2.25)), and we have temporarily introduced the notation

$$
\left[\frac{\partial \widetilde{e}_{0,\theta}}{\partial \boldsymbol{\theta}}\right] := \begin{bmatrix} \dfrac{\partial e^{d+1}_{0,\theta}}{\partial \theta^{d+1}} & \cdots & \dfrac{\partial e^{d+1}_{0,\theta}}{\partial \theta^{d+n}} \\ \vdots & & \vdots \\ \dfrac{\partial e^{d+2n}_{0,\theta}}{\partial \theta^{d+1}} & \cdots & \dfrac{\partial e^{d+2n}_{0,\theta}}{\partial \theta^{d+n}} \end{bmatrix} \in \mathrm{M}_{2n,n}(\mathbb{R}) \ .
$$

The expression for $\partial_\omega \widetilde{Y}_{0,\theta}$ then becomes (cf. (3.38))

$$\partial_\omega \widetilde{Y}_{0,\theta} = \left[\frac{\partial \widetilde{V}_{\lambda_0}}{\partial \widetilde{x}}\right]_{K_0(\theta)} \widetilde{Y}_\theta + 2\widetilde{J}^{-1}_{K_0(\theta)} \widetilde{\Pi}_{0,\theta} \left[\frac{\partial \widetilde{V}_{\lambda_0}}{\partial \widetilde{x}}\right]^{\mathrm{sym}}_{K_0(\theta)} \widetilde{X}_\theta R_{0,\theta}$$
$$+ \widetilde{J}^{-1}_{K_0(\theta)}\, D\widetilde{J}_{K_0(\theta)}\, e_{0,\theta}\, \widetilde{Y}_0 - \widetilde{J}^{-1}_{K_0(\theta)} \left[\frac{\partial \widetilde{e}_{0,\theta}}{\partial \boldsymbol{\theta}}\right] R_{0,\theta}$$
$$+ \widetilde{Y}_{0,\theta} \left\{ \left[\frac{\partial \widetilde{e}_{0,\theta}}{\partial \boldsymbol{\theta}}\right]^\top \widetilde{X}_{0,\theta} + \widetilde{X}^\top_{0,\theta} \left[\frac{\partial \widetilde{e}_{0,\theta}}{\partial \boldsymbol{\theta}}\right] \right\} R_{0,\theta} \ ,$$

which implies (cf. (3.39))

$$\left(DV_{\lambda_0, K_0(\theta)} - \partial_\omega\right) M_{0,\theta} = \begin{bmatrix} \left(DV_{\lambda_0, K_0(\theta)} - \partial_\omega\right) DK_{0,\theta} & \left(DV_{\lambda_0, K_0(\theta)} - \partial_\omega\right) \begin{bmatrix} 0 \\ \widetilde{Y}_{0,\theta} \end{bmatrix} \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 & \left[\frac{\partial \underline{V}_{\lambda_0}}{\partial \widetilde{x}}\right]_{K_0(\theta)} \widetilde{Y}_{0,\theta} \\ 0 & 0 & -2\widetilde{J}^{-1}_{K_0(\theta)} \widetilde{\Pi}_{0,\theta} \left[\frac{\partial \widetilde{V}_{\lambda_0}}{\partial \widetilde{x}}\right]^{\mathrm{sym}}_{K_0(\theta)} \widetilde{X}_\theta R_\theta \end{bmatrix} + \begin{bmatrix} De_{0,\theta} & \begin{bmatrix} 0 \\ \mathscr{E}[e_0](\theta) \end{bmatrix} \end{bmatrix} . \tag{4.20}$$

Here we have set

$$\mathscr{E}[e_0](\theta) := \widetilde{J}^{-1}_{K_0(\theta)}\, D\widetilde{J}_{K_0(\theta)}\, e_{0,\theta}\, \widetilde{Y}_{0,\theta} - \widetilde{J}^{-1}_{K_0(\theta)} \left[\frac{\partial \widetilde{e}_{0,\theta}}{\partial \boldsymbol{\theta}}\right] R_{0,\theta}$$
$$+ \widetilde{Y}_{0,\theta} \left(\left[\frac{\partial \widetilde{e}_{0,\theta}}{\partial \boldsymbol{\theta}}\right]^\top \widetilde{X}_{0,\theta} + \widetilde{X}^\top_{0,\theta} \left[\frac{\partial \widetilde{e}_{0,\theta}}{\partial \boldsymbol{\theta}}\right] \right) R_{0,\theta} \in \mathrm{M}_{2n,n}(\mathbb{R}) \ , \tag{4.21}$$

and $\widetilde{\Pi}_{0,\theta}$ is defined as in (3.37), but with $\widetilde{X}_\theta$ and R_θ replaced by $\widetilde{X}_{0,\theta}$ and $R_{0,\theta}$, respectively.

The first matrix in the right-hand side of (4.20) is the "big" contribution (i.e., the one that does not vanish when e_0 is set to 0), and the second one is the "error" which becomes zero when e_0 is identically 0.

To rewrite the coefficient of $\xi_{0,\theta}$ in (4.19) in a simple form, we want that

$$(DV_{\lambda_0,K_0(\theta)} - \partial_\omega)M_{0,\theta} = M_{0,\theta}\,(C_{0,\theta} + B_{0,\theta}) \ , \tag{4.22}$$

where $C_{0,\theta}$ has the form

$$C_{0,\theta} = \begin{bmatrix} 0 & 0 & T_{0,\theta} \\ 0 & 0 & S_{0,\theta} \\ 0 & 0 & 0 \end{bmatrix} \ , \tag{4.23}$$

and $B_{0,\theta}$ is a "small" matrix, i.e., a matrix that vanishes if e_0 becomes identically zero. The equations (4.22) and (4.23) should be compared with (3.41) and (3.42).

Now we will compute explicit expressions for $T_{0,\theta}$, $S_{0,\theta}$ and $B_{0,\theta}$. To take care of the "big" terms in the right-hand sides of (4.20) and (4.22), we equate the product $M_{0,\theta}C_{0,\theta}$ (with $M_{0,\theta}$ and $C_{0,\theta}$ given by (4.16) and (4.23)) with the "big" term in the right-hand side of (4.20):

$$\begin{bmatrix} \underline{Z}_{0,\theta} & \underline{X}_{0,\theta} & 0 \\ \widetilde{Z}_{0,\theta} & \widetilde{X}_{0,\theta} & \widetilde{Y}_{0,\theta} \end{bmatrix} \begin{bmatrix} 0 & 0 & T_{0,\theta} \\ 0 & 0 & S_{0,\theta} \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \left[\dfrac{\partial \underline{V}_{\lambda_0}}{\partial \widetilde{x}}\right]_{K_0(\theta)} \widetilde{Y}_{0,\theta} \\ 0 & 0 & -2\widetilde{J}^{-1}_{K_0(\theta)}\widetilde{\Pi}_{0,\theta}\left[\dfrac{\partial \widetilde{V}_{\lambda_0}}{\partial \widetilde{x}}\right]^{\mathrm{sym}}_{K_0(\theta)} \widetilde{X}_\theta R_\theta \end{bmatrix} .$$

This gives us the system

$$\begin{aligned} \underline{Z}_{0,\theta}T_{0,\theta} + \underline{X}_{0,\theta}S_{0,\theta} &= \left[\frac{\partial \underline{V}_{\lambda_0}}{\partial \widetilde{x}}\right]_{K_0(\theta)} \widetilde{J}^{-1}_{K_0(\theta)}\widetilde{X}_{0,\theta}R_{0,\theta} \\ \widetilde{Z}_{0,\theta}T_{0,\theta} + \widetilde{X}_{0,\theta}S_{0,\theta} &= -2\widetilde{J}^{-1}_{K_0(\theta)}\widetilde{\Pi}_{0,\theta}\left[\frac{\partial \widetilde{V}_{\lambda_0}}{\partial \widetilde{x}}\right]^{\mathrm{sym}}_{K_0(\theta)} \widetilde{X}_{0,\theta}R_{0,\theta} \ , \end{aligned}$$

whose solution is (cf. (3.48) and (3.50))

$$T_{0,\theta} = \underline{\mathscr{Z}}_{0,\theta}^{-1} \left(\left[\frac{\partial \underline{V}_{\lambda_0}}{\partial \widetilde{x}} \right]_{K_0(\theta)} \widetilde{Y}_{0,\theta} - 2 \underline{X}_{0,\theta} \widetilde{Y}_{0,\theta}^{\top} \widetilde{\Pi}_{0,\theta} \left[\frac{\partial \widetilde{V}_{\lambda_0}}{\partial \widetilde{x}} \right]_{K_0(\theta)}^{\mathrm{sym}} \widetilde{X}_{0,\theta} R_{0,\theta} \right) ,$$

$$\begin{aligned} S_{0,\theta} = {} & -\widetilde{Y}_{0,\theta}^{\top} \widetilde{J}_{K_0(\theta)} \widetilde{Z}_{0,\theta} \underline{\mathscr{Z}}_{0,\theta}^{-1} \left[\frac{\partial \underline{V}_{\lambda_0}}{\partial \widetilde{x}} \right]_{K_0(\theta)} \widetilde{Y}_{0,\theta} \\ & + 2 \left(\mathbb{I}_n - \widetilde{Y}_{0,\theta}^{\top} \widetilde{J}_{K_0(\theta)} \widetilde{Z}_{0,\theta} \underline{\mathscr{Z}}_{0,\theta}^{-1} \underline{X}_{0,\theta} \right) \widetilde{Y}_{0,\theta}^{\top} \widetilde{\Pi}_{0,\theta} \left[\frac{\partial \widetilde{V}_{\lambda_0}}{\partial \widetilde{x}} \right]_{K_0(\theta)}^{\mathrm{sym}} \widetilde{X}_{0,\theta} R_{0,\theta} , \end{aligned} \tag{4.24}$$

where, similarly to (3.49), we have set

$$\underline{\mathscr{Z}}_{0,\theta} := \underline{Z}_{0,\theta} + \underline{X}_{0,\theta} \widetilde{Y}_{0,\theta}^{\top} \widetilde{J}_{K_0(\theta)} \widetilde{Z}_{0,\theta} \in \mathrm{M}_{d,d}(\mathbb{R}) . \tag{4.25}$$

Using (4.22), we can rewrite equation (4.19) in the form

$$(C_{0,\theta} + B_{0,\theta})\, \xi_{0,\theta} - \partial_\omega \xi_{0,\theta} = -M_{0,\theta}^{-1} \left(e_{0,\theta} + \left[\frac{\partial V_\lambda}{\partial \lambda} \right]_{\lambda_0, K_0(\theta)} \varepsilon_0 \right) \tag{4.26}$$

where $C_{0,\theta}$ is given by (4.23) and (4.24), and $B_{0,\theta}$ is a "small" matrix, given (according to (4.20)) by

$$B_{0,\theta} = M_{0,\theta}^{-1} \begin{bmatrix} D e_{0,\theta} & \begin{bmatrix} 0 \\ \mathscr{E}[e_0](\theta) \end{bmatrix} \end{bmatrix} , \tag{4.27}$$

with $\mathscr{E}[e_0](\theta)$ defined in (4.21).

4.3.3 Invertibility issues

Since the rank of the $(2n \times 2n)$-matrix $[\widetilde{X}_{0,\theta}\ \widetilde{Y}_{0,\theta}]$ is maximal and the matrix $\widetilde{J}_{K_0(\theta)}$ is non-degenerate, the matrix

$$Q_{0,\theta} := \begin{bmatrix} \mathbb{I}_d & 0 \\ 0 & \widetilde{X}_{0,\theta}^\top \\ 0 & \widetilde{Y}_{0,\theta}^\top \end{bmatrix} \begin{bmatrix} \mathbb{I}_d & 0 \\ 0 & \widetilde{J}_{K_0(\theta)} \end{bmatrix} = \begin{bmatrix} \mathbb{I}_d & 0 \\ 0 & \widetilde{X}_{0,\theta}^\top \widetilde{J}_{K_0(\theta)} \\ 0 & \widetilde{Y}_{0,\theta}^\top \widetilde{J}_{K_0(\theta)} \end{bmatrix} ,$$

defined similarly to Q_θ (3.55), is non-degenerate.

In the spirit of (3.56), define the matrix

$$W_{0,\theta} := \begin{bmatrix} \underline{Z}_{0,\theta} & \underline{X}_{0,\theta} & 0 \\ 0 & 0 & \mathbb{I}_n \\ \widetilde{Y}_{0,\theta}^\top \widetilde{J}_{K_0(\theta)} \widetilde{Z}_{0,\theta} & -\mathbb{I}_n & \widetilde{Y}_{0,\theta}^\top \widetilde{J}_{K_0(\theta)} \widetilde{Y}_{0,\theta} \end{bmatrix} . \tag{4.28}$$

The motivation in defining $W_{0,\theta}$ is that it is approximately equal to $Q_{0,\theta} M_{0,\theta}$—indeed,

$$Q_{0,\theta} M_{0,\theta} = \begin{bmatrix} \underline{Z}_{0,\theta} & \underline{X}_{0,\theta} & 0 \\ \widetilde{X}_{0,\theta}^\top \widetilde{J}_{K_0(\theta)} \widetilde{Z}_{0,\theta} & \widetilde{X}_{0,\theta}^\top \widetilde{J}_{K_0(\theta)} \widetilde{X}_{0,\theta} & \mathbb{I}_n \\ \widetilde{Y}_{0,\theta}^\top \widetilde{J}_{K_0(\theta)} \widetilde{Z}_{0,\theta} & -\mathbb{I}_n & \widetilde{X}_{0,\theta}^\top \widetilde{J}_{K_0(\theta)} \widetilde{Y}_{0,\theta} \end{bmatrix} ,$$

so that

$$Q_{0,\theta} M_{0,\theta} = W_{0,\theta} + P_{0,\theta} , \tag{4.29}$$

where the matrix

$$P_{0,\theta} := \begin{bmatrix} 0 & 0 & 0 \\ \widetilde{X}_{0,\theta}^\top \widetilde{J}_{K_0(\theta)} \widetilde{Z}_{0,\theta} & \widetilde{X}_{0,\theta}^\top \widetilde{J}_{K_0(\theta)} \widetilde{X}_{0,\theta} & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{4.30}$$

is small; if K_0 were a true solution, $P_{0,\theta}$ would be zero (recall (3.57)).

Lemma 4.3. *Assume that the hypotheses of Lemma 4.2 hold. Then there exists a constant C depending on d, n, σ, ρ, $\|DK_0\|_\rho$, $|V_{\lambda_0}|_{C^1,\mathcal{B}_r}$, and $|J|_{C^1,\mathcal{B}_r}$, such that for every δ satisfying*

$$0 < \delta < \frac{\rho}{2} \ ,$$

the following bound holds:

$$\|W_0^{-1}\, P_0\|_{\rho-2\delta} \leq C\gamma^{-1}\delta^{-(\sigma+1)}\|e_0\|_\rho \ . \tag{4.31}$$

Proof. Recalling the bound (4.5) on the norm of the pull-back $L_{0,\theta}$ (4.4) of the presymplectic form Ω to the torus $\mathcal{K}_0 = K_0(\mathbb{T}^{d+n})$, we obtain

$$\begin{aligned}
\|W_0^{-1} P_0\|_{\rho-2\delta} &\leq C\, \|P_0\|_{\rho-2\delta} \\
&\leq C_1\, \|\widetilde{X}_0^\top\, (\widetilde{J} \circ K_0)\, \widetilde{Z}_0\|_{\rho-2\delta} + C_2\, \|\widetilde{X}_0^\top\, (\widetilde{J} \circ K_0)\, \widetilde{X}_0\|_{\rho-2\delta} \\
&\leq C\, \|DK_0^\top\, (\widetilde{J} \circ K_0)\, DK_0\|_{\rho-2\delta} \\
&= C\, \|L_0\|_{\rho-2\delta} \\
&\leq C\gamma^{-1}\delta^{-(\sigma+1)}\, \|e_0\|_\rho \ .
\end{aligned}$$

$\square$

The approximate factorization (4.29) can be used to write the inverse matrix $M_{0,\theta}^{-1}$ in a convenient form, and Lemma 4.3 yields some useful bounds:

Lemma 4.4. *Assume that the hypotheses of Lemma 4.2 hold. Assume that*

$$0 < \delta < \frac{\rho}{2}$$

and the error e_0 satisfies the bound

$$C\gamma^{-1}\delta^{-(\sigma+1)}\|e_0\|_\rho \le \frac{1}{2} \ , \tag{4.32}$$

where C is the same constant as in the right-hand side of (4.31)*.*

Then the matrix $M_{0,\theta}$ is invertible and its inverse can be written in the form

$$M_{0,\theta}^{-1} = W_{0,\theta}^{-1} Q_{0,\theta} + M_{\mathrm{E},\theta} \ , \tag{4.33}$$

where the error term, $M_{\mathrm{E},\theta}$, is given by

$$M_{\mathrm{E},\theta} = -\left(\mathbb{I}_{d+2n} + W_{0,\theta}^{-1} P_{0,\theta}\right)^{-1} W_{0,\theta}^{-1} P_{0,\theta} W_{0,\theta}^{-1} Q_{0,\theta} \ , \tag{4.34}$$

and satisfies the bound

$$\|M_{\mathrm{E}}\|_{\rho-2\delta} \le C' \gamma^{-1} \delta^{-(\sigma+1)} \|e_0\|_\rho \ ; \tag{4.35}$$

here C' is a constant that depends on the same parameters as the constant C in the right-hand side of the bound (4.31)*.*

Proof. From (4.29) written in the form

$$Q_{0,\theta}M_{0,\theta} = W_{0,\theta} + P_{0,\theta} = W_{0,\theta}\left(\mathbb{I}_{d+2n} + W_{0,\theta}^{-1}P_{0,\theta}\right) \ ,$$

we obtain

$$M_{0,\theta}^{-1} = \left(\mathbb{I}_{d+2n} + W_{0,\theta}^{-1}P_{0,\theta}\right)^{-1} W_{0,\theta}^{-1}Q_{0,\theta} \ .$$

From the definition (4.33) of $M_{\mathrm{E},\theta}$ we easily obtain the explicit expression (4.34):

$$\begin{aligned}
M_{\mathrm{E},\theta} &= M_{0,\theta}^{-1} - W_{0,\theta}^{-1}Q_{0,\theta} \\
&= \left(\mathbb{I}_{d+2n} + W_{0,\theta}^{-1}P_{0,\theta}\right)^{-1} W_{0,\theta}^{-1}Q_{0,\theta} - W_{0,\theta}^{-1}Q_{0,\theta} \\
&= \left(\mathbb{I}_{d+2n} + W_{0,\theta}^{-1}P_{0,\theta}\right)^{-1} \left\{\mathbb{I}_{d+2n} - \left(\mathbb{I}_{d+2n} + W_{0,\theta}^{-1}P_{0,\theta}\right)\right\} W_{0,\theta}^{-1}Q_{0,\theta} \\
&= -\left(\mathbb{I}_{d+2n} + W_{0,\theta}^{-1}P_{0,\theta}\right)^{-1} W_{0,\theta}^{-1}P_{0,\theta}W_{0,\theta}^{-1}Q_{0,\theta} \ .
\end{aligned}$$

The bound (4.35) is a direct consequence of (4.31). □

Remark 3. Because of the special form of $W_{0,\theta}$ (4.28), its inverse also has a special form, namely

$$W_{0,\theta}^{-1} =$$

$$\begin{bmatrix}
\underline{\mathscr{Z}}_{\theta}^{-1} & -\underline{\mathscr{Z}}_{0,\theta}^{-1}\,\underline{X}_{0,\theta}\,(W_{0,\theta})^{3}{}_{3} & \underline{\mathscr{Z}}_{0,\theta}^{-1}\,\underline{X}_{0,\theta} \\
(W_{0,\theta})^{3}{}_{1}\,\underline{\mathscr{Z}}_{0,\theta}^{-1} & \left\{\mathbb{I}_n - (W_{0,\theta})^{3}{}_{1}\,\underline{\mathscr{Z}}_{0,\theta}^{-1}\,\underline{X}_{0,\theta}\right\}(W_{0,\theta})^{3}{}_{3} & -\mathbb{I}_n + (W_{0,\theta})^{3}{}_{1}\,\underline{\mathscr{Z}}_{0,\theta}^{-1}\,\underline{X}_{0,\theta} \\
0 & \mathbb{I}_n & 0
\end{bmatrix}$$

where $(W_{0,\theta})^{3}{}_{1} := \widetilde{Y}_{0,\theta}^{\top}\widetilde{J}_{K_0(\theta)}\widetilde{Z}_{0,\theta}$ and $(W_{0,\theta})^{3}{}_{3} := \widetilde{Y}_{0,\theta}^{\top}\widetilde{J}_{K_0(\theta)}\widetilde{Y}_{0,\theta}$ are the corresponding matrix elements of $W_{0,\theta}$, and $\underline{\mathscr{Z}}_{0,\theta}$ is defined in (4.25).

4.3.4 Bounds on the "small" parts

Recall that, in order to find an approximate solution of the linearized equation (4.12), we changed the variable $\Delta_{0,\theta}$ to $\xi_{0,\theta}$ by (4.13) to transform it to the form (4.19). Then we rewrote the coefficient of $\xi_{0,\theta}$ in (4.19) as a sum of a "big" part, $C_{0,\theta}$ (given by (4.23) and (4.24)), and a "small" part, $B_{0,\theta}$, given by (4.27). In the Proposition below we give bounds on the "small" terms in (4.19).

Proposition 4.5. *Let $K_0 \in \mathcal{W}_\rho$ and the error e_0 be defined by* (4.1)*. Let the pair (λ_0, K_0) be non-degenerate for the family V_λ of presymplectic analytic vector fields in the sense of Definition 4.6. If the error e_0 satisfies* (4.32)*, then the change of variables* (4.13) *transforms equation* (4.12) *to*

$$\begin{aligned}\left(\begin{bmatrix} 0 & 0 & T_{0,\theta} \\ 0 & 0 & S_{0,\theta} \\ 0 & 0 & 0 \end{bmatrix} + B_{0,\theta}\right)\xi_{0,\theta} - \partial_\omega \xi_{0,\theta} &= -M_{0,\theta}^{-1}\left(e_{0,\theta} + \partial_\lambda V_{\lambda_0,K_0(\theta)}\,\varepsilon_0\right) \\ &= -W_{0,\theta}^{-1}\,Q_{0,\theta}\,e_{0,\theta} - W_{0,\theta}^{-1}\,Q_{0,\theta}\,\left[\frac{\partial V_\lambda}{\partial \lambda}\right]_{\lambda_0,K_0(\theta)}\varepsilon_0 \\ &\quad - M_{\mathrm{E},\theta}\,e_{0,\theta} - M_{\mathrm{E},\theta}\,\left[\frac{\partial V_\lambda}{\partial \lambda}\right]_{\lambda_0,K_0(\theta)}\varepsilon_0\ ,\end{aligned} \tag{4.36}$$

where $B_{0,\theta}$ is defined by (4.27)*, M_{E} is given by* (4.34) *(and satisfies* (4.33)*).*

Furthermore, the following bounds hold:

$$\|B_0\|_{\rho-2\delta} \leq C\delta^{-1} \|e_0\|_\rho \ , \tag{4.37}$$

$$\|M_{\mathrm{E}}\, e_0\|_{\rho-2\delta} \leq C\gamma^{-1}\delta^{-(\sigma+1)} \|e_0\|_\rho^2 \ , \tag{4.38}$$

$$\left\| M_{\mathrm{E}} \left[\frac{\partial V_\lambda}{\partial \lambda}\bigg|_{\lambda_0} \circ K_0 \right] \varepsilon_0 \right\|_{\rho-2\delta} \leq C\gamma^{-1}\delta^{-(\sigma+1)} \left\| \frac{\partial V_\lambda}{\partial \lambda}\bigg|_{\lambda_0} \circ K_0 \right\| \|e_0\|_\rho |\varepsilon_0| \ . \tag{4.39}$$

Proof. Equation (4.36) follows directly from the variational equation written in the form (4.26), where $C_{0,\theta}$ is given by (4.23) and (4.24), $B_{0,\theta}$ is given by (4.27), and the representation (4.33) of $M_{0,\theta}^{-1}$ is used. So we only need to derive the bounds (4.37), (4.38), and (4.39).

Combining (4.27) and (4.33), we obtain

$$B_{0,\theta} = \left(W_{0,\theta}^{-1} Q_{0,\theta} + M_{\mathrm{E},\theta}\right) \begin{bmatrix} De_{0,\theta} & \begin{bmatrix} 0 \\ \mathscr{E}[e_0](\theta) \end{bmatrix} \end{bmatrix} .$$

From the definition (4.21) of $\mathscr{E}[e_0](\theta)$ and the Cauchy bound (2.15),

$$\|\mathscr{E}[e_0]\|_{\rho-2\delta} \leq C_1 \|e_0\|_{\rho-2\delta} + C_2\delta^{-1} \|e_0\|_{\rho-\delta} \leq C\delta^{-1} \|e_0\|_{\rho-\delta} .$$

This, together with the bound (4.35) on M_{E}, yields (4.37):

$$\begin{aligned} \|B_0\|_{\rho-2\delta} &\leq \left(\left\|W_0^{-1}Q_0\right\|_{\rho-2\delta} + \|M_{\mathrm{E}}\|_{\rho-2\delta}\right) \left(\|De_0\|_{\rho-2\delta} + \|\mathscr{E}[e_0]\|_{\rho-2\delta}\right) \\ &\leq \left(C_1 + C_2\gamma^{-1}\delta^{-(\sigma+1)}\|e_0\|_\rho\right) \gamma^{-1} \|e_0\|_{\rho-\delta} \\ &\leq C\gamma^{-1} \|e_0\|_\rho \ . \end{aligned}$$

The bounds (4.38) and (4.39) are direct consequences of (4.35). $\square$

To use Newton method for finding $\xi_{0,\theta}$, it is enough to solve (4.36) retaining only the "big" terms, i.e., ignoring all terms that are of higher order with respect to the norm of the error e_0. As we will show below (see (4.50)), the term ε_0 is also of order of the norm of e_0. Proposition 4.5 allows us to keep only the leading terms in (4.36) and write

$$\begin{bmatrix} 0 & 0 & T_{0,\theta} \\ 0 & 0 & S_{0,\theta} \\ 0 & 0 & 0 \end{bmatrix} \xi_{0,\theta} - \partial_\omega \xi_{0,\theta} = -W_{0,\theta}^{-1}\, Q_{0,\theta}\, e_{0,\theta} - \Lambda_{0,\theta}\, \varepsilon_0 \ , \tag{4.40}$$

where we have set

$$\Lambda_{0,\theta} := W_{0,\theta}^{-1}\, Q_{0,\theta} \left[\frac{\partial V_\lambda}{\partial \lambda}\right]_{\lambda_0, K_0(\theta)} \ .$$

Introducing the notation

$$\xi_{0,\theta} =: \begin{bmatrix} \xi^{\mathrm{z}}_{0,\theta} \\ \xi^{\mathrm{x}}_{0,\theta} \\ \xi^{\mathrm{y}}_{0,\theta} \end{bmatrix} \ , \qquad \xi^{\mathrm{z}}_{0,\theta} \in \mathrm{M}_{d,1}(\mathbb{R}) \ , \quad \xi^{\mathrm{x}}_{0,\theta},\, \xi^{\mathrm{y}}_{0,\theta} \in \mathrm{M}_{n,1}(\mathbb{R}) \ ,$$

we write (4.40) in the form

$$\partial_\omega \begin{bmatrix} \xi^{\mathrm{z}}_{0,\theta} \\ \xi^{\mathrm{x}}_{0,\theta} \\ \xi^{\mathrm{y}}_{0,\theta} \end{bmatrix} = W_{0,\theta}^{-1}\, Q_{0,\theta}\, e_{0,\theta} + \Lambda_{0,\theta}\, \varepsilon_0 + \begin{bmatrix} T_{0,\theta}\, \xi^{\mathrm{y}}_{0,\theta} \\ S_{0,\theta}\, \xi^{\mathrm{y}}_{0,\theta} \\ 0 \end{bmatrix} \ . \tag{4.41}$$

Equation (4.41) has a solution if and only if the average over $\mathbb{T}^{d+n}$ of its right-hand side is 0:

$$\operatorname{avg}\left(W_0^{-1}\, Q_0\, e_0\right) + \operatorname{avg}\left(\Lambda_0\right) \varepsilon_0 + \begin{bmatrix} \operatorname{avg}\left(T_0\, \xi_0^{\mathrm{y}}\right) \\ \\ \operatorname{avg}\left(S_0\, \xi_0^{\mathrm{y}}\right) \\ \\ 0 \end{bmatrix} = 0\ . \tag{4.42}$$

To satisfy the condition (4.42), we could have determined ε_0 from this equation and then substitute this value for ε_0 into (4.41) to find the solution ξ_0. The problem with this strategy is that we still do not know ξ_0^{y}. What saves the strategy is the observation that right-hand side of the last n equations of the system (4.41) does not involve ξ_0^{y}, so that the last n equations of (4.41) have the form

$$\partial_\omega \xi_{0,\theta}^{\mathrm{y}} = \left(W_{0,\theta}^{-1}\, Q_{0,\theta}\, e_{0,\theta} + \Lambda_{0,\theta}\, \varepsilon_0\right)^{\mathrm{y}}\ , \tag{4.43}$$

for which the solvability condition is

$$\operatorname{avg}\left(W_0^{-1}\, Q_0\, e_0\right) + \operatorname{avg}\left(\Lambda_0\right) \varepsilon_0 = 0\ . \tag{4.44}$$

In order to guarantee that we can solve (4.44) for ε_0, we have to require that the matrix multiplying ε_0 is non-degenerate. To this end, we give the following

Definition 4.6. *The pair (λ_0, K_0) is said to be* ***non-degenerate*** *for a $(d+2n)$-parameter family of vector fields V_λ if the matrix*

$$\Lambda_{0,\theta} := W_{0,\theta}^{-1} Q_{0,\theta} \left[\frac{\partial V_\lambda}{\partial \lambda}\right]_{\lambda_0, K_0(\theta)} \in \mathrm{M}_{d+2n,d+2n}(\mathbb{R})$$

has a non-singular average:

$$\operatorname{rank}\operatorname{avg}\left(\Lambda_0\right) = d + 2n\ , \qquad \operatorname{avg}\left(\Lambda_0\right) = \int_{\mathbb{T}^{d+n}} \Lambda_{0,\theta}\, \mathrm{d}\theta^1 \cdots \mathrm{d}\theta^{d+n}\ . \tag{4.45}$$

We assume that the non-degeneracy condition (4.45) is satisfied, and set ε_0 to be equal to the preliminary value

$$\varepsilon_0^{\mathrm{prelim}} := -\left\{\operatorname{avg}(\Lambda_0)\right\}^{-1} \operatorname{avg}\left(W_0^{-1}\, Q_0\, e_0\right)\ , \tag{4.46}$$

which is of order of the norm of the error:

$$|\varepsilon_0^{\mathrm{prelim}}| \le C \operatorname{avg}(e_0) \le C\, \|e_0\|_\rho\ . \tag{4.47}$$

This choice of ε_0 guarantees the existence of a solution ξ_0^{y} of (4.43) that satisfies the bound

$$\begin{aligned} \|\xi_0^{\mathrm{y}}\|_{\rho-\delta} &\le C\gamma^{-1}\delta^{-\sigma} \left\| W_0^{-1}\, Q_0\, e_0 + \Lambda_0 \varepsilon_0^{\mathrm{prelim}} \right\|_\rho \\ &\le C\gamma^{-1}\delta^{-\sigma}\, \|e_0\|_\rho \end{aligned} \tag{4.48}$$

thanks to the Rüssmann's inequality (2.16) and the bound (4.47).

Having found ξ_0^{y} from solving (4.43), we redefine ε_0 as

$$\varepsilon_0 := -\left\{\operatorname{avg}(\Lambda_0)\right\}^{-1} \left(\operatorname{avg}\left(W_0^{-1}\, Q_0\, e_0\right) + \begin{bmatrix} \operatorname{avg}(T_0\, \xi_0^{\mathrm{y}}) \\ \operatorname{avg}(S_0\, \xi_0^{\mathrm{y}}) \\ 0 \end{bmatrix} \right) \tag{4.49}$$

to satisfy condition (4.42). Note that, although the value of ε_0 from (4.49) differs from the preliminary choice (4.46), the new value of ε_0 will still satisfy the solvability condition (4.44) so that (4.43) will still be solvable. Thanks to the bound

(4.48), the updated value of ε_0 satisfies

$$\begin{aligned}|\varepsilon_0| &\le C\left(\|e_0\|_{\rho-\delta} + \|\xi_0^{\mathrm{y}}\|_{\rho-\delta}\right)\\ &\le C\left(\|e_0\|_\rho + C\gamma^{-1}\delta^{-\sigma}\|e_0\|_\rho\right)\\ &\le C\gamma^{-1}\delta^{-\sigma}\|e_0\|_\rho\ .\end{aligned} \tag{4.50}$$

With the new value of ε_0 from (4.49), we solve (4.41) to find ξ_0 which, according to the Rüssmann's inequality (2.16) and the bound (4.50), satisfies

$$\begin{aligned}\|\xi_0\|_{\rho-2\delta} &\le C\gamma^{-1}\delta^{-\sigma}\left(\left\|W_0^{-1}\,Q_0\,e_0 + \Lambda_0\varepsilon_0\right\|_{\rho-\delta} + C\,\|\xi_0^{\mathrm{y}}\|_{\rho-\delta}\right)\\ &\le C\gamma^{-1}\delta^{-\sigma}\left(\|e_0\|_\rho + |\varepsilon_0| + \gamma^{-1}\delta^{-\sigma}\|e_0\|_\rho\right)\\ &\le C\gamma^{-2}\delta^{-2\sigma}\|e_0\|_\rho\ .\end{aligned} \tag{4.51}$$

We have just proved the following lemma:

Lemma 4.7. *Assume the hypotheses of Proposition 4.5. Then there exist a function, ξ_0, and a parmeter, ε_0, that solve the reduced linear equation* (4.40) *and satisfy the bounds* (4.50) *and* (4.51).

Chapter 5

Newton Method

In this chapter we will present estimates for the j^{th} step of the iterative scheme and show that the Newton Method generates a Cauchy sequence of approximate solutions in a Banach space which converges to a true solution.

5.1 Improved-Step Estimates

Lemma 5.1. *Assume that* $(\lambda_j\,, K_j)$ *is an approximate solution with the same assumptions as Proposition* 4.7 *such that the following holds:*

$$r_j := \|K_j - K_0\|_{\rho_j} < r\,. \tag{5.1}$$

If $\|e_j\|_{\rho_j}$ *is small enough such that Proposition* 4.7 *applies, then there exist a function* Δ_j *and a parameter* $\varepsilon_j \in \mathbb{R}^{d+2n}$ *such that*

$$\begin{aligned} \|\Delta_j\|_{\rho_j-2\delta_j} &\leq c_j \gamma^{-2} \delta_j^{-2\sigma} \|e_j\|_{\rho_j} \\ \|D\Delta_j\|_{\rho_j-3\delta_j} &\leq c_j \gamma^{-2} \delta_j^{-(2\sigma+1)} \|e_j\|_{\rho_j} \\ |\varepsilon_j| &\leq c_j \left|\operatorname{avg}(\Lambda_j)^{-1}\right| \|e_j\|_{\rho_j} \ , \end{aligned} \tag{5.2}$$

where c_j *is a constant that depends on* $n, d, r, \rho, |V_{\lambda_j}|_{\mathcal{C}^2,\mathcal{B}_r}, \|DK_j\|_{\rho_j}, \|R_j\|_{\rho_j}$, *and* $\left\|\frac{\partial V_\lambda}{\partial \lambda}\right\|_{\rho_j}$ *Additionally, if*

$$r_j + c_j \gamma^{-2} \delta_j^{-(2\sigma-1)} \|e_j\|_{\rho_j} < r \ , \tag{5.3}$$

then

$$\|e_{j+1}\|_{\rho_{j+1}} \leq c_j \gamma^{-4} \delta_j^{-4\sigma} \|e_j\|_{\rho_j}^2 \ . \tag{5.4}$$

Proof. The inequalities (5.2) follow directly from Proposition 4.5, the Cauchy inequality, and the fact that $\Delta_j = M_j \, \xi_j$.

We have defined $K_{j+1,\theta} = K_{j,\theta} + \Delta_{j,\theta}$. So

$$\begin{aligned} \|K_{j+1} - K_0\|_{\rho_{j+1}-2\delta_{j+1}} &= \|K_j + \Delta_j - K_0\|_{\rho_{j+1}-2\delta_{j+1}} \\ &\leq \|K_j - K_0\|_{\rho_j} + \|\Delta_j\|_{\rho_j-2\delta_j} \\ &\leq r_j + c_j \gamma^{-2} \delta_j^{-(2\sigma+1)} \|e_j\|_{\rho_j} \ , \end{aligned}$$

which is smaller than r by the assumption (5.3). This means that $K_{j+1} \in \mathcal{B}_r$, that is, our new approximate solution stays within the neighborhood where V is holomorphically extened.

To see that (5.4) is true, recall from equation (4.10) that $\xi_j = M_j^{-1} \Delta_j$ was found by solving (4.40). Thus,

$$
\begin{aligned}
DV_{\lambda_j,K_j(\theta)}\,\Delta_{j,\theta} - \partial_\omega \Delta_{j,\theta} + \left[\frac{\partial V_\lambda}{\partial \lambda}\right]_{\lambda_j,K_j(\theta)} \varepsilon_j + e_j \\
= M_{j,\theta}\left(B_{j,\theta}\xi_{j,\theta} + M_{E,j,\theta}e_{j,\theta} + M_{E,j,\theta}\left[\left.\frac{\partial V_\lambda}{\partial \lambda}\right|_{\lambda_j} \circ K_j\right]\right),
\end{aligned}
$$

and each term on the right hand side is quadratically small from Proposition 4.5. This gives us the bound

$$
\left\| DV_{\lambda_j,K_j}\,\Delta_j - \partial_\omega \Delta_j + \left[\frac{\partial V_\lambda}{\partial \lambda}\right]_{\lambda_j,K_j} \varepsilon_j + e_j \right\|_{\rho_j - 2\delta_j} \le C\gamma^{-3}\delta^{-(3\sigma+1)} \left\| e_j \right\|_{\rho_j}^2 . \quad (5.5)
$$

Finally, recalling the Taylor expansion of $e_{j+1,\theta}$ as given in (4.10) we see that the size of the remainder term is on the order of $\|\Delta_j\|_{\rho_j-2\delta_j}^2$. Thus we get the estimate (5.4). □

5.2 Non-degeneracy conditions

This section will show that if the error is small enough and some invertibility conditions are met, then they will also be met at the subsequent (improved) step.

Lemma 5.2. *Assume the setup of Lemma* 5.1. *If*

$$
c_j \gamma^{-2} \delta_j^{-(\sigma+1)} \left\| e_j \right\|_{\rho_j} \le \frac{1}{2}, \quad (5.6)
$$

then the following are true:

1. *If $\widetilde{X}_j^\top \widetilde{X}_j$ is invertible, then $\widetilde{X}_{j+1}^\top \widetilde{X}_{j+1}$ is invertible.*

2. If W_j is invertible, then W_{j+1} is invertible.

3. If $\operatorname{avg}(\Lambda_j)$ is invertible, then $\operatorname{avg}(\Lambda_{j+1})$ is invertible.

Proof. Define

$$D\Delta_j := \begin{bmatrix} \underline{\Delta}_{j,\underline{x}} & \underline{\Delta}_{j,\widetilde{x}} \\ \widetilde{\Delta}_{j,\underline{x}} & \widetilde{\Delta}_{j,\widetilde{x}} \end{bmatrix},$$

then

$$DK_{j+1} = D\left(K_j + \Delta_j\right) = DK_j + D\Delta_j = \begin{bmatrix} \underline{Z}_j & \underline{X}_j \\ \widetilde{Z}_j & \widetilde{X}_j \end{bmatrix} + \begin{bmatrix} \underline{\Delta}_{j,\underline{x}} & \underline{\Delta}_{j,\widetilde{x}} \\ \widetilde{\Delta}_{j,\underline{x}} & \widetilde{\Delta}_{j,\widetilde{x}} \end{bmatrix}$$

and

$$\widetilde{X}_{j+1} = \widetilde{X}_j + \widetilde{\Delta}_{j,\widetilde{x}} .$$

Therefore

$$\begin{aligned} \widetilde{X}_{j+1}^{\top}\widetilde{X}_{j+1} &= \left(\widetilde{X}_j + \widetilde{\Delta}_{j,\widetilde{x}}\right)^{\top}\left(\widetilde{X}_j + \widetilde{\Delta}_{j,\widetilde{x}}\right) \\ &= \widetilde{X}_j^{\top}\widetilde{X}_j + \widetilde{X}_j^{\top}\widetilde{\Delta}_{j,\widetilde{x}} + \widetilde{\Delta}_{j,\widetilde{x}}^{\top}\widetilde{X}_j + \widetilde{\Delta}_{j,\widetilde{x}}^{\top}\widetilde{\Delta}_{j,\widetilde{x}} \\ &= \widetilde{X}_j^{\top}\widetilde{X}_j + P_j , \end{aligned}$$

where $P_j := \widetilde{X}_j^{\top}\widetilde{\Delta}_{j,\widetilde{x}} + \widetilde{\Delta}_{j,\widetilde{x}}^{\top}\widetilde{X}_j + \widetilde{\Delta}_{j,\widetilde{x}}^{\top}\widetilde{\Delta}_{j,\widetilde{x}}$.

The first term, $\widetilde{X}_j^{\top}\widetilde{X}_j$, is invertible by assumption, and the three terms that make up P_j are all bounded by Δ_j, which is bounded by the size of the error. Thus, $\mathbb{I}_n + \left(\widetilde{X}_j^{\top}\widetilde{X}_j\right)^{-1}P_j$ is invertible by the Neumann series, so

$$\widetilde{X}_{j+1}^{\top}\widetilde{X}_{j+1} = \left(\widetilde{X}_j^{\top}\widetilde{X}_j\right)\left(\mathbb{I}_n + \left(\widetilde{X}_j^{\top}\widetilde{X}_j\right)^{-1}P_j\right)$$

is also invertible.

The other terms follow similar arguments and use the Neumann Series to establish their invertibility. The key point is that we are only changing the term by a small amount and invertiblity is an open condition. □

5.3 Convergence to a true solution

In this section, we will show how close the initial approximation has to be for our method to be iterated indefinitely and to converge to a true solution. Also, we will get a bound on the difference between the true solution and the initial approximation.

Lemma 5.3. *Let* $\{c_j\}_{j\geq 0}$ *be the sequence of constants given above. Then for* $0 < \delta_0 < \min(\rho_0/12\,,1)$ *define:*

$$\begin{aligned}
\delta_j &:= \delta_0 2^{-j}\ , \\
\rho_j &:= \rho_{j-1} - 6\delta_{j-1}\ , \\
\rho_\infty &:= \lim_{j\to\infty} \rho_j\ , \\
r_j &:= \|K_j - K_0\|_{\rho_j}\ , \\
K_\infty &:= \lim_{j\to\infty} K_j\ .
\end{aligned}$$

Then there exists a constant $C > 0$ *depending on* d, n, $|V_\lambda|_{\mathcal{C}^2,\mathcal{B}_r}$, $|J_0|_{\mathcal{C}^1,\mathcal{B}_r}$, $\|DK_0\|_{\rho_0}$, *and* $|\{\mathrm{avg}(\Lambda_0)\}^{-1}|$ *such that if* $\|e_0\|_{\rho_0}$ *satisfies the conditions*

$$C 2^{4\sigma}\gamma^{-4}\delta_0^{-4\sigma}\,\|e_0\|_{\rho_0} \leq \frac{1}{2}\ , \tag{5.7}$$

$$C\left(1 + \frac{2^{4\sigma}}{2^{2\sigma}-1}\right)\gamma^{-2}\delta_0^{-2\sigma}\,\|e_0\|_{\rho_0} < r\ , \tag{5.8}$$

then the Newton Method can be successively iterated and will converge to a true solution, $(\lambda_\infty, K_\infty)$.

Furthermore, the following bound holds:

$$\|K_\infty - K_0\|_{\rho_0 - 6\delta_0} \leq \left(\frac{2^{2\sigma}}{2^{2\sigma}-1}\right) c\gamma^{-2}\delta_0^{-2\sigma} \|e_0\|_{\rho_0}\,. \tag{5.9}$$

Proof. The proof of this lemma is completely standard in KAM and follows the details of [26]. The main point is that if (5.7) and (5.8) are true, then for all $j \geq 0$, we have the following:

$$r_j + C\gamma^{-2}\delta_j^{-2\sigma} \|e_j\|_{\rho_j} < r \tag{5.10}$$

$$C\gamma^{-2}\delta_j^{-(\sigma+1)} \|e_j\|_{\rho_j} \leq \frac{1}{2}\,. \tag{5.11}$$

This ensures that at each step, the improved approximate torus, K_j, stays within $\mathcal{B}_r$ and that the conditions are right for M_j to be inverted so that we may solve the reduced linear equation and then change variables to get the update function Δ_j. This part of the Newton Method would be the same as in [26] with little modifiction, so in order to save the reader the headache of reading through a long and tedious induction proof, we willl present only a few of the inequatlities and give some general comments.

Notice that

$$
\begin{aligned}
\|e_j\|_{\rho_j} &\le C\gamma^{-4}\delta_{j-1}^{-4\sigma}\|e_{j-1}\|_{\rho_{j-1}}^2 \\
&\le C\gamma^{-4}\delta_{j-1}^{-4\sigma}\left(C\gamma^{-4}\delta_{j-2}^{-4\sigma}\|e_{j-2}\|_{\rho_{j-2}}^2\right)^2 \\
&\le \dots \\
&\le \left(C\gamma^{-4}\delta_0^{-4\sigma}\right)^{1+2+\dots+2^{j-1}}\left(\delta_{j-1}\delta_{j-2}^2\dots\delta_0^{2^{j-1}}\right)\|e_0\|_{\rho_0}^{2^j} \\
&= \left(C\gamma^{-4}\delta_0^{-4\sigma}\right)^{2^j-1}\left(2^{4\sigma}\right)^{2^0(j-1)+2^1(j-2)+\dots+2^{j-2}(1)}\|e_0\|_{\rho_0}^{2^j} \\
&\le \left(C\gamma^{-4}\delta_0^{-4\sigma}\right)^{2^j-1}2^{4\sigma(2^j-j)}\|e_0\|_{\rho_0}^{2^j} \\
&\le \left(C\gamma^{-4}\delta_0^{-4\sigma}2^{4\sigma}\|e_0\|_{\rho_0}\right)^{2^j-1}2^{-4\sigma(j-1)}\|e_0\|_{\rho_0} \\
&\le \kappa^{2^j-1}2^{-4\sigma(j-1)}\|e_0\|_{\rho}\,,
\end{aligned}
$$

where

$$
\kappa := C\gamma^{-4}\delta_0^{-4\sigma}2^{4\sigma}\|e_0\|_{\rho_0}\,. \tag{5.12}
$$

Thus, we have

$$
\begin{aligned}
c_j\gamma^{-2}\delta_j^{-(2\sigma+1)}\|e_j\|_{\rho_j} &\le \left(C\gamma^{-2}\delta_0^{-(2\sigma+1)}\|e_0\|_{\rho_0}\right)2^{4\sigma}\kappa^{(2^j-1)}2^{-j(2\sigma-1)} \\
&\le \kappa\gamma^2\delta_0^{2\sigma-1}\kappa^{2^j-1}2^{-j(2\sigma-1)} \le \frac{1}{2}\,,
\end{aligned}
$$

Also, we can see that

$$
\begin{aligned}
r_j &= \|K_j - K_0\|_{\rho_j} \\
&\le \|K_{j-1} + \Delta_{j-1} - K_0\|_{\rho_{j-1}} \\
&\le \|K_{j-1} - K_0\|_{\rho_{j-1}} + c\gamma^{-2}\delta_{j-1}^{-2\sigma} \|e_{j-1}\|_{\rho_{j-1}} \\
&= r_{j-1} + c\gamma^{-2}\delta_{j-1}^{-2\sigma} \|e_{j-1}\|_{\rho_{j-1}} \\
&\le \dots \\
&\le c\gamma^{-2}\delta_0^{-2\sigma} \|e_0\|_{\rho_0} + c\gamma^{-2} \sum_{m=1}^{j-1} \delta_m^{-2\sigma} \|e_m\|_{\rho_m} \\
&\le c\gamma^{-2}\delta_0^{-2\sigma} \|e_0\|_{\rho_0} + c\gamma^{-2}\delta_0^{-2\sigma} \kappa \|e_0\|_{\rho_0} \sum_{m=1}^{j-1} 2^{2m\sigma} 2^{-4\sigma(m-1)} \\
&\le c\gamma^{-2}\delta_0^{-2\sigma} \|e_0\|_{\rho_0} \left(1 + \kappa 2^{2\sigma} \sum_{m=0}^{\infty} 2^{-2m\sigma}\right) \\
&= c\gamma^{-2}\delta_0^{-2\sigma} \|e_0\|_{\rho_0} \left(1 + \kappa \frac{2^{4\sigma}}{2^{2\sigma} - 1}\right) \,,
\end{aligned}
$$

which is less than r because $\kappa < \frac{1}{2}$.

Thus the sequence $\{K_j\}_{j \ge 0}$ forms a Cauchy sequence in a Banach space and will converge to a true soluition $K_\infty = \lim_{j \to \infty} K_j$.

Finally, recalling (5.2), we obtain

$$
\begin{aligned}
\|K_\infty - K_0\|_{\rho_\infty} &\le \sum_{j=0}^{\infty} \|K_0 + \Delta_0 + \Delta_1 + \Delta_2 + \cdots - K_0\|_{\rho_\infty} \\
&\le \sum_{j=0}^{\infty} \|\Delta_j\|_{\rho_\infty} \\
&\le \sum_{j=0}^{\infty} \|\Delta_j\|_{\rho_j} \\
&\le \sum_{j=0}^{\infty} C\gamma^{-2}\delta_j^{-2\sigma} \|e_j\|_{\rho_j} \\
&\le \sum_{j=0}^{\infty} C\gamma^{-2}\delta_0^{-2\sigma} 2^{2j\sigma} \kappa^{2^j-1} 2^{-4\sigma(j-1)} \|e_0\|_{\rho} \\
&\le \sum_{j=0}^{\infty} C\kappa^{2^j-1} 2^{-2j\sigma+4\sigma} \gamma^{-2}\delta_0^{-2\sigma} \|e_0\|_{\rho} \ .
\end{aligned}
$$

□

Bibliography

[1] R. Abraham and J. E. Marsden. *Foundations of Mechanics*. Benjamin/Cummings, Reading, Mass., second edition, 1978.

[2] R. Abraham, J. E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications*, volume 75 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition, 1988.

[3] L. Abrunheiro, M. Camarinha, and J. Clemente-Gallardo. Cubic polynomials on Lie groups: reduction of the Hamiltonian system. *J. Phys. A*, 44(35):355203, 16, 2011. Corrigendum: 46(18):189501, 2, 2013.

[4] H. N. Alishah and R. de la Llave. Tracing KAM tori in presymplectic dynamical systems. *J. Dynam. Differential Equations*, 24(4):685–711, 2012.

[5] J. L. Anderson and P. G. Bergmann. Constraints in covariant field theories. *Physical Rev. (2)*, 83:1018–1025, 1951.

[6] V. I. Arnol′d. Proof of a theorem of A. N. Kolmogorov on the preservation of conditionally periodic motions under a small perturbation of the Hamiltonian. *Uspehi Mat. Nauk*, 18(5 (113)):13–40, 1963.

[7] V. I. Arnold, V. V. Kozlov, and A. I. Neishtadt. *Mathematical Aspects of Classical and Celestial Mechanics*, volume 3 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, third edition, 2006. [Dynamical systems. III].

[8] D. Aubin and A. Dahan Dalmedico. Writing the history of dynamical systems and chaos: longue durée and revolution, disciplines and cultures. *Historia Math.*, 29(3):273–339, 2002.

[9] M. Barbero-Liñán and M. C. Muñoz-Lecanda. Geometric approach to Pontryagin's maximum principle. *Acta Appl. Math.*, 108(2):429–485, 2009.

[10] C. Batlle, J. Gomis, J. M. Pons, and N. Roman. Lagrangian and Hamiltonian constraints. *Lett. Math. Phys.*, 13(1):17–23, 1987.

[11] C. Batlle, J. Gomis, J. M. Pons, and N. Román-Roy. Equivalence between the Lagrangian and Hamiltonian formalism for constrained systems. *J. Math. Phys.*, 27(12):2953–2962, 1986.

[12] G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn. A proof of Kolmogorov's theorem on invariant tori using canonical transformations defined by the Lie method. *Nuovo Cimento B (11)*, 79(2):201–223, 1984.

[13] P. G. Bergmann and I. Goldberg. Dirac bracket transformations in phase space. *Phys. Rev. (2)*, 98:531–538, 1955.

[14] H. Broer and M. Sevryuk. KAM theory: quasi-periodicity in dynamical systems. In *Handbook of Dynamical Systems. Vol. 3*, pages 249–344. Elsevier/North-Holland, Amsterdam, 2010.

[15] R. Calleja and R. de la Llave. A numerically accessible criterion for the breakdown of quasi-periodic solutions and its rigorous justification. *Nonlinearity*, 23(9):2029–2058, 2010.

[16] R. C. Calleja, A. Celletti, and R. de la Llave. A KAM theory for conformally symplectic systems: efficient algorithms and their validation. *J. Differential Equations*, 255(5):978–1049, 2013.

[17] J. F. Cariñena, J. Gomis, L. A. Ibort, and N. Román. Canonical transformations theory for presymplectic systems. *J. Math. Phys.*, 26(8):1961–1969, 1985.

[18] J. F. Cariñena and L. A. Ibort. Geometric theory of the equivalence of Lagrangians for constrained systems. *J. Phys. A*, 18(17):3335–3341, 1985.

[19] J. F. Cariñena, C. Lopez, and N. Román-Roy. Geometric study of the connection between the Lagrangian and Hamiltonian constraints. *J. Geom. Phys.*, 4(3):315–334, 1987.

[20] J. F. Cariñena and J. Nasarre. On the symplectic structures arising in geometric optics. *Fortschr. Phys.*, 44(3):181–198, 1996.

[21] J. F. Cariñena and J. Nasarre. Presymplectic geometry and Fermat's principle for anisotropic media. *J. Phys. A*, 29(8):1695–1702, 1996.

[22] L. Chierchia. KAM lectures. In *Dynamical Systems. Part I*, Pubbl. Cent. Ric. Mat. Ennio Giorgi, pages 1–55. Scuola Norm. Sup., Pisa, 2003. Hamiltonian systems and celestial mechanics, Selected papers from the Research Trimester held in Pisa, February 4–April 26, 2002.

[23] L. Chierchia. A. N. Kolmogorov's 1954 paper on nearly-integrable Hamiltonian systems. *Regul. Chaotic Dyn.*, 13(2):130–139, 2008.

[24] L. Chierchia. Kolmogorov-Arnold-Moser (KAM) theory. In R. A. Meyers, editor, *Encyclopedia of Complexity and Systems Science*, pages 5064–5091. Springer, 2009.

[25] R. H. Cushman and L. M. Bates. *Global Aspects of Classical Integrable Systems.* Birkhäuser/Springer, Basel, second edition, 2015.

[26] R. de la Llave, A. González, À. Jorba, and J. Villanueva. KAM theory without action-angle variables. *Nonlinearity*, 18(2):855–895, 2005.

[27] M. de León and D. Martín de Diego. Almost product structures and Poisson reduction of presymplectic systems. *Extracta Math.*, 10(1):37–45, 1995.

[28] P. A. M. Dirac. Generalized Hamiltonian dynamics. *Canadian J. Math.*, 2:129–148, 1950.

[29] P. A. M. Dirac. The Hamiltonian form of field dynamics. *Canadian J. Math.*, 3:1–23, 1951.

[30] P. A. M. Dirac. Generalized Hamiltonian dynamics. *Proc. Roy. Soc. London. Ser. A*, 246:326–332, 1958.

[31] P. A. M. Dirac. *Lectures on Quantum Mechanics*, volume 2 of *Belfer Graduate School of Science Monographs Series.* Belfer Graduate School of Science, New York, 1964.

[32] H. S. Dumas. *The KAM Story: A Friendly Introduction to the Content, History, and Significance of Classical Kolmogorov-Arnold-Moser Theory.* World Scientific, New Jersey, 2014.

[33] C. Duval. Finsler spinoptics. *Comm. Math. Phys.*, 283(3):701–727, 2008.

[34] C. Duval. Polarized spinoptics and symplectic physics. *Preprint*, 2013. arXiv:1312.4486 [math-ph].

[35] J. Earman. Tracking down gauge: an ode to the constrained Hamiltonian formalism. In K. Brading and E. Castellani, editors, *Symmetries in Physics: Philosophical Reflections*, pages 140–162. Cambridge University Press, 2003.

[36] A. Echeverría-Enríquez, J. Marín-Solano, M. C. Muñoz-Lecanda, and N. Román-Roy. Geometric reduction in optimal control theory with symmetries. *Rep. Math. Phys.*, 52(1):89–113, 2003.

[37] A. Echeverría-Enríquez, M. C. Muñoz-Lecanda, and N. Román-Roy. Reduction of presymplectic manifolds with symmetry. *Rev. Math. Phys.*, 11(10):1209–1247, 1999.

[38] E. Fontich, R. de la Llave, and Y. Sire. Construction of invariant whiskered tori by a parameterization method. I. Maps and flows in finite dimensions. *J. Differential Equations*, 246(8):3136–3213, 2009.

[39] E. Fontich, R. de la Llave, and Y. Sire. A method for the study of whiskered quasi-periodic and almost-periodic solutions in finite and infinite dimensional Hamiltonian systems. *Electron. Res. Announc. Math. Sci.*, 16:9–22, 2009.

[40] A. M. Fox and J. D. Meiss. Critical invariant circles in asymmetric and multiharmonic generalized standard maps. *Commun. Nonlinear Sci. Numer. Simul.*, 19(4):1004–1026, 2014.

[41] A. González-Enríquez, A. Haro, and R. de la Llave. Singularity theory for non-twist KAM tori. *Mem. Amer. Math. Soc.*, 227(1067):vi+115, 2014.

[42] M. Gotay. *Presymplectic Manifolds, Geometric Constraint Theory and Dirac-Bergmann Theory of Constraints*. PhD thesis, University of Maryland, 1979.

[43] M. J. Gotay and J. M. Nester. Presymplectic Lagrangian systems. I. The constraint algorithm and the equivalence theorem. *Ann. Inst. H. Poincaré Sect. A (N.S.)*, 30(2):129–142, 1979.

[44] M. J. Gotay and J. M. Nester. Generalized constraint algorithm and special presymplectic manifolds. In *Geometric Methods in Mathematical Physics (Proc. NSF-CBMS Conf., Univ. Lowell, Lowell, Mass., 1979)*, volume 775 of *Lecture Notes in Math.*, pages 78–104. Springer, Berlin, 1980.

[45] M. J. Gotay and J. M. Nester. Presymplectic Lagrangian systems. II. The second-order equation problem. *Ann. Inst. H. Poincaré Sect. A (N.S.)*, 32(1):1–13, 1980.

[46] M. J. Gotay, J. M. Nester, and G. Hinds. Presymplectic manifolds and the Dirac-Bergmann theory of constraints. *J. Math. Phys.*, 19(11):2388–2399, 1978.

[47] A. Hanson, T. Regge, and C. Teitelboim. *Constrained Hamiltonian Systems*. Accademia Nazionale dei Lincei, Roma, 1976.

[48] À. Haro, M. Canadell, A. Luque, J. M. Mondelo, and J.-L. Figueras. *The Parameterization Method for Invariant Manifolds: From Rigorous Results to Effective Computations*, volume 195 of *Applied Mathematical Sciences*. Springer-Verlag, 2016.

[49] M. Henneaux and C. Teitelboim. *Quantization of Gauge Systems*. Princeton University Press, Princeton, NJ, 1992.

[50] G. Huguet, R. de la Llave, and Y. Sire. Computation of whiskered invariant tori and their associated manifolds: new fast algorithms. *Discrete Contin. Dyn. Syst.*, 32(4):1309–1353, 2012.

[51] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*. Cambridge University Press, Cambridge, 1995.

[52] A. N. Kolmogorov. On conservation of conditionally periodic motions for a small change in Hamilton's function. *Dokl. Akad. Nauk SSSR (N.S.)*, 98:527–530, 1954.

[53] A. N. Kolmogorov. Théorie générale des systèmes dynamiques et mécanique classique. In *Proceedings of the International Congress of Mathematicians, Amsterdam, 1954, Vol. 1*, pages 315–333. Erven P. Noordhoff N.V., Groningen; North-Holland, Amsterdam, 1957.

[54] A. N. Kolmogorov. Preservation of conditionally periodic movements with small change in the Hamilton function. In *Stochastic Behavior in Classical and Quantum Hamiltonian Systems (Volta Memorial Conf., Como, 1977)*, volume 93 of *Lecture Notes in Phys.*, pages 51–56. Springer, Berlin-New York, 1979.

[55] P. Libermann and C.-M. Marle. *Symplectic Geometry and Analytical Mechanics*. D. Reidel, Dordrecht, 1987.

[56] R. de la Llave. A tutorial on KAM theory. In *Smooth Ergodic Theory and Its Applications (Seattle, WA, 1999)*, volume 69 of *Proc. Sympos. Pure Math.*, pages 175–292. Amer. Math. Soc., Providence, RI, 2001.

[57] A. Luque and J. Villanueva. A KAM theorem without action-angle variables for elliptic lower dimensional tori. *Nonlinearity*, 24(4):1033–1080, 2011.

[58] J. Moser. A rapidly convergent iteration method and non-linear differential equations. II. *Ann. Scuola Norm. Sup. Pisa (3)*, 20:499–535, 1966.

[59] J. Moser. A rapidly convergent iteration method and non-linear partial differential equations. I. *Ann. Scuola Norm. Sup. Pisa (3)*, 20:265–315, 1966.

[60] J. Moser. *Stable and Random Motions in Dynamical Systems*. Princeton University Press, Princeton, N. J.; University of Tokyo Press, Tokyo, 1973. With special emphasis on celestial mechanics, Hermann Weyl Lectures, the Institute for Advanced Study, Princeton, N. J, Annals of Mathematics Studies, No. 77.

[61] E. Newman and P. G. Bergmann. Lagrangians linear in the "velocities". *Phys. Rev. (2)*, 99:587–592, 1955.

[62] I. Niven. *Diophantine Approximations*. John Wiley & Sons, New York-London, 1963.

[63] T. Noda and M. Oda. Geometry of moment maps and reductions for presymplectic manifolds. *Georgian Math. J.*, 16(2):343–352, 2009.

[64] J. Pöschel. A lecture on the classical KAM theorem. In *Smooth Ergodic Theory and Its Applications (Seattle, WA, 1999)*, volume 69 of *Proc. Sympos. Pure Math.*, pages 707–732. Amer. Math. Soc., Providence, RI, 2001.

[65] H. J. Rothe and K. D. Rothe. *Classical and Quantum Dynamics of Constrained Hamiltonian Systems*, volume 81 of *World Scientific Lecture Notes in Physics*. World Scientific, Hackensack, NJ, 2010.

[66] G. Rudolph and M. Schmidt. *Differential Geometry and Mathematical Physics, I: Manifolds, Lie Grouips and Hamiltonian Systems*. Springer, 2013.

[67] H. Rüssmann. On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus. In *Dynamical Systems, Theory and Applications (Rencontres, Battelle Res. Inst., Seattle, Wash., 1974)*, pages 598–624. Lecture Notes in Phys., Vol. 38. Springer, Berlin, 1975.

[68] D. A. Salamon. The Kolmogorov-Arnold-Moser theorem. *Math. Phys. Electron. J.*, 10:Paper 3, 37 pp. (electronic), 2004.

[69] E. C. G. Sudarshan and N. Mukunda. *Classical Dynamics: A Modern Perspective*. Wiley-Interscience, New York-London-Sydney, 1974.

[70] K. Sundermeyer. *Constrained Dynamics*, volume 169 of *Lecture Notes in Physics*. Springer-Verlag, Berlin-New York, 1982.

[71] K. Sundermeyer. *Symmetries in Fundamental Physics*, volume 176 of *Fundamental Theories of Physics*. Springer, Cham, 2014.

[72] W. Thirring. *A Course in Mathematical Physics. Vol. I. Classical Dynamical Systems*. Springer-Verlag, New York-Vienna, 1978.

[73] F. W. Warner. *Foundations of Differentiable Manifolds and Lie Groups*. Scott, Foresman and Co., Glenview, Ill.-London, 1971.

[74] A. Wipf. Hamilton's formalism for systems with constraints. In *Canonical Gravity: From Classical to Quantum (Bad Honnef, 1993)*, volume 434 of *Lecture Notes in Phys.*, pages 22–58. Springer, Berlin, 1994.