Introduction to KAM theory

August 31, 2023

Machiel Kruger, 6224059

Master thesis

Universiteit Utrecht

Abstract

This thesis aims to provide an introduction to modern KAM theory.

This objective is accomplished by introducing the fundamental ideas of KAM theory, including its background, commonly used methods, and the general structure of a KAM proof. Additionally, the thesis discusses parameter reduction and the generalizations of the Diophantine conditions. The culmination of the thesis is a complete proof of the classical KAM theorem, but the primary focus remains pedagogical and expository throughout.

It is important to note that while the thesis covers a broad range of topics, it does not aim to provide a systematic survey of all the methods found in KAM theory. Instead, the emphasis is on presenting the traditional KAM method and its key ideas in a clear and accessible manner.

Contents

Acknowledgements

To begin, I would like to express my sincere gratitude to my primary supervisor,, Heinz Hanßmann, for providing me with numerous helpful references, insightful comments, and for his patience in dealing with the numerous spelling and grammar errors this text once had.

Secondly, I would like to thank the board of examiners for giving me the opportunity to thoroughly study this subject by gratuitously extending the time frame.

Lastly, I would also like to express my appreciation to my parents and friends who provided invaluable support in completing this project within the given time constraints.

Introduction

This thesis serves as an introductory to Kolmogorov-Arnol'd-Moser theory, named after its founders, and commonly abbreviated as KAM theory today.

The theory is fundamental in the study of quasi-periodic motions in dynamical systems across various contexts. Such motions densely fill invariant tori, on which the dynamics are aperiodic. More precisely, the frequency ratio on these tori is irrational, resulting in an orbit that densely fills the invariant torus. Such tori are referred to as quasi-periodic invariant tori. KAM theory is concerned with the behavior of these quasi-periodic invariant tori under small perturbations and specifically addresses whether or not these tori persist. Contrary to earlier beliefs[1], KAM theory has proven the persistence of most of these tori under suitably small perturbations. Thus, it proves that quasi-periodicity can survive small perturbations.

This was a major breakthrough, and warrants recognition as a landmark theorem within the field of dynamics.

KAM theory has since been applied to a wide array of situations involving quasi-periodic invariant tori and small perturbations. Most notably, KAM theory serves as the foundation for quasi-periodic bifurcation theory, which studies the bifurcation of quasi-periodic invariant tori and still is an active area of research.

However, while of major importance within mathematics, KAM theory has become less popular outside of it due to its limited[2] applicability to practical problems.

Nevertheless, KAM theory has had major impact on the development of statistical mechanics, most notably on the ergodic hypothesis. It disproved the hypothesis by demonstrating a type of stability in non-integrable systems, which by the hypothesis were thought to be purely ergodic.

KAM theory has also been successfully applied to very specific n-body problems, where it has proven the long-term stability of several 3-body scenarios. However, KAM theory typically requires very small thresholds on the acceptable size of the perturbations under which many of the quasi-periodic invariant tori are proven to persist[3]. Currently, it is quite impossible to measure the mass of astronomical bodies to the precision required by KAM theory thresholds. Therefore KAM theory has had limited applicability to the study of the stability of (non-idealised) solar systems.

Intent

As with most classical subjects, this is not the first introduction that has been written on the subject. There are already quite a few well-written tutorials and lecture notes on KAM theory, such as those in [19, 28, 29, 56, 73], and the monographs [17, 18, 20, 38]. However, most of the well-known tutorials and introductions on KAM theory were written in the 1990s and 2000s, and since then, the proofs and methods have been refined. Hence, I believe there is still some room for improvement. However, due to the quality of the already available introductory texts, this by no means is a complete replacement, and I encourage interested readers to explore the available literature.

Most[4] of the pedagogic proofs and demonstrations cover only the methods required for the demonstrated proof itself. Although these provide a very good starting point, the numerous technicalities involved in KAM theory can obscure the bigger picture. Due to size constraints, the general methods are usually wordlessly applied.

The tutorials and introductions that do not include full KAM proofs, but mostly reference the monographs and the aforementioned pedagogic proofs, quickly become overwhelming with the amount of literature that is available on KAM theory.

Lastly, the published literature itself is also far ahead of most examples and tutorials, making the step from tutorials to modern studies quite challenging.

This thesis aims to explain KAM theory in more manageable pieces while still maintaining an overview of the literature. It also includes a demonstration proof, which is more verbose than is usually the case in KAM theory. Specifically, I aim to demonstrate a few of the KAM theory methods used in quasi-periodic bifurcation theory, as this was the path that led me to KAM theory.

Due to the extensive literature on KAM theory, it is not feasible to offer a comprehensive survey in this thesis. Thus, I acknowledge that this introduction is not exhaustive, and further reading is necessary to fully comprehend the theory. Below, I provide an outline of the thesis, which clearly presents its contents.

[1] Notably, the non-persistence was falsely conjectured by Poincaré.

[2] KAM theory usually requires very small perturbations, which are difficult (or even impossible) to measure or limit in real life situations.

[3] It is possible for single tori to survive under perturbations well outside this bound.

[4] That I personally have read for the study for this thesis.

Outline

Introduction

The introduction begins with the definition of quasi-periodicity and briefly covers perturbation theory, which is the study of the effect of small perturbations on integrable systems - a fundamental problem in dynamics. KAM theory is a newer tool in the study of this problem and, being essentially a perturbation theory, shares many of its motivations and applications with earlier perturbation theories.

However, the methods to achieve this result are more advanced and require more conditions. The implicit function theorem breaks down fast when considering quasi-periodic orbits and conjugation theorems suffer from the small "divisors"[5] or more accurately small denominators.

The introduction discusses the similarities and differences between KAM theory and traditional perturbation theory, as well as provides a brief historical overview and references to the most significant variations of the KAM theorem.

Part I: Small denominator problems

In part I, small denominator problems are explored in conjunction with Whitney differentiability.

The part begins with a discussion of the complex linearization problem, which is one of the most well-known small denominator problems, next to classical KAM theory, and closely related to KAM theory as some methods are shared between them.

Afterwards, the Diophantine conditions for KAM theory for the flow of vector fields are defined. These conditions define the set of Diophantine frequencies, and the properties of this set are proven and their implications are considered. Additionally, a small subsection on more generalized conditions on the frequencies is included. Closely related to the structure of the set of Diophantine frequencies is the regularity in the Whitney sense of functions defined on such sets. This regularity is discussed in terms of Whitney differentiability.

Lastly, a demonstration of a small denominator problem and Whitney differentiability is given in the form of a solution to the one-bite small denominator problem.

Part II: Geometry

Part II focuses on the geometry surrounding KAM theory, beginning with an overview of the four main geometric settings. These settings are defined by a geometric structure on the manifold, which all vector fields on the manifold are expected to preserve.

The symplectic form, which defines the structure of the Hamiltonian context, is explained in more detail, and methods for constructing symplectomorphisms are discussed. In particular, the Lie algebra method is discussed, which relies on the fact that vector fields that preserve an admissible[6] geometric structure form Lie algebras, with structure-preserving diffeomorphisms as the associated Lie group. This enables an easier calculation of structure-preserving conjugations.

Next, the part proceeds with the definition of integrability in KAM theory and briefly discusses Floquet tori, along with the full Lagrangian context. The parameters, frequency map, and non-degeneracy conditions of KAM theory are also explored. Finally, an overview of the general KAM method for constructing conjugations is provided.

Part III: KAM proof

In part III, the theory and methods from previous parts are applied to prove a frequency decoupled KAM theorem in the classical Hamiltonian setting[7]. Due to the length of the proof, the lemmas used in the proof are omitted to maintain the flow of the proof. Finally, the frequency decoupled theorem is used to prove the classical KAM theorem.

Appendix

The appendix consists of two sections. The first section contains the proofs of all the lemmas used in the KAM proof. The second section contains several tools, such as Cauchy's estimate and the Paley-Wiener lemma, along with some technical lemmas used in the first appendix section and the KAM proof. There is also a quick subsection on differential geometry.

[5] The term "divisors" is the result of a mistranslation.
[6] Not all geometric structure form a Lie algebra.
[7] Also known as the Lagrangian or maximal Hamiltonian context.

1.1 Quasi-periodicity

One of the most common methods of studying dynamical systems is to analyze the properties of specific dynamical features found within them. This approach simplifies the study of systems by breaking them down into their constituent parts and is widely applied both within and outside the realm of mathematics. The simplest features studied are fixed points and periodic orbits, and these have been well understood for some time. More recently, complex features such as homoclinic and heteroclinic orbits have also been studied.

A common question asked about these dynamical features is whether they are structurally stable, i.e. do they persist under small perturbations?

This line of inquiry has resulted in a wide array of theories concerning the persistence of many features, and KAM theory is one such persistence theory. The dynamical features considered by KAM theory are invariant quasi-periodic tori. KAM theory gives insight in which quasi-periodic tori survive perturbation, this is not generally true and requires a specific set of conditions.

Definitions

Let X be a smooth vector field on a manifold M. We assume that M contains an n-torus T that is invariant under the flow of X on M.

Definition 1.1.1 (Parallel flow)**.** If there exists a diffeomorphism from T to the standard n-torus $\mathbb{T}^n$ such that the restriction of X to T, $X|_T$, transforms to a constant vector field on $\mathbb{T}^n$, given by $\sum_{i=0}^{n} \omega_i \partial/\partial x_i$, then we say X induces *parallel flow* on T. The vector $\omega \in \mathbb{R}^n$ is known as the *frequency vector*[8].

Invariant tori with more complicated nonlinear flow do exist. In this thesis (and in KAM theory), we only consider the simple case, where we only consider tori with parallel flow and small perturbations of such tori. The dynamics on invariant tori with parallel flow are highly dependent on the number-theoretical properties of the frequency vector.

Definition 1.1.2 (Quasi-periodic tori)**.** The parallel flow of an invariant torus is called *quasi-periodic* or *non-resonant* if all frequencies ω_i are rationally independent, equivalently if $\langle \omega, k \rangle \neq 0$ for all $k \in \mathbb{Z}^n$. A torus equipped with quasi-periodic parallel flow is known as a *quasi-periodic torus*.

If there does exists an integer $k \in \mathbb{Z}^n$ such that $\langle \omega, k \rangle = 0$, then the parallel flow is called a *resonant*, and a torus equipped with resonant parallel flow is known as a *resonant* torus. If two or more frequencies are rationally dependent, we refer to them as *resonant* frequencies or *resonances* in the torus. In this thesis, we only consider the simple case where the frequencies are either all non-resonant or all resonant. Cases where not all frequencies are rationally independent foliate the invariant torus into many lower-dimensional tori. Such cases are much more difficult to handle, and fall outside the scope of this thesis.

Under this assumption, every orbit on a resonant torus is periodic and therefore a closed invariant subset of the torus. Therefore, each orbit can be studied individually as a distinct dynamical feature. Since individual periodic orbits are easier to study, resonant tori are often studied as a family of periodic orbits. The orbits on a quasi-periodic torus are still invariant by definition of an orbit. However, unlike the resonant case, quasi-periodic orbits are not closed as they do not repeat themselves exactly after a finite time. As a result, quasi-periodic orbits cannot be studied as dynamical features since they do not form closed invariant subsets.

Therefore, an invariant quasi-periodic torus cannot be nicely split into constituent parts and should be studied as a whole. Consequently, the quasi-periodic tori are coherent or indivisible dynamical features, in a similar vein as fixed points and periodic orbits.

Not all quasi-periodic tori survive under small perturbations, as with frequencies too "close" to resonant frequencies difficulties with small denominators arise. Persistence has only been proven for a class of strongly non-resonant tori. The most common method to define this class is to require each non-resonant frequency to satisfy Diophantine conditions.

Definition 1.1.3 (Diophantine tori)**.** An invariant torus is called *Diophantine* and *strongly non-resonant* if for some constants $\tau > 0$ and $\gamma > 0$ the frequency vector ω satisfies

$$\langle \omega, k \rangle \geq \frac{\gamma}{|k|^\tau} \quad \text{for all} \quad k \in \mathbb{Z}^n \setminus \{0\}. \tag{1.1}$$

A frequency vector that satisfies (1.1) is referred to as a *Diophantine* frequency vector.

KAM theory proves the persistence of Diophantine tori, under sufficiently small perturbations and an additional non-degeneracy condition.

[8] Uniquely determined up to transformation by $\omega \mapsto A\omega$, with $A \in GL(n, \mathbb{Z})$.

1.2 KAM theory as a perturbation theory

Before we consider KAM theory and its additional difficulties and conditions, we first provide an introduction to the more general perturbation theory. Perturbation theory is a rich field, encompassing numerous theorems and methods applicable to a variety of contexts. We begin by discussing the fundamental elements of perturbation theory, followed by its general applications.

1.2.1 Anatomy of a perturbation theory

Every perturbation theory is set within a specific class of dynamical systems. A dynamical system consists of a manifold M, which can have several additional properties that the dynamical system may preserve. These include a symplectic or volume-preserving form, or any other enforced symmetry, such as reversibility. The dynamics can be either continuous or discrete. The dynamical setting is usually called the *setting* or, in the case of KAM theory, the *context* of the perturbation theory.

On these manifolds, we define a specific unperturbed system. Usually, the unperturbed systems are chosen to be easy to work with and, more importantly, contain some feature or property of interest. Examples include linear systems, normal form truncations, integrable systems, and generally any system with well-known and well-studied dynamics. Many perturbation theories are only valid for a certain class of unperturbed systems, and such a class is sometimes defined as a class of *integrable* systems for the perturbation theory in question.

Lastly, we consider the perturbations of this system. These are usually restricted to a specific type that, for example, always preserve the additional properties associated with the setting, such as the symplectic form. Perturbations could be required to come in a specific form relevant to its application, such as a restriction to certain coordinates. The size of the perturbation, in a suitable topology, is also very commonly restricted. This is generally what is meant by a suitably "small" perturbation.

A perturbation theorem proves, in a certain setting, that the well-understood dynamics of a class of integrable systems are preserved under a set of perturbations, where usually the integrable systems are required to meet additional non-degeneracy conditions. If features are proven to persist under small perturbations, it is possible to define a continuous map from the unperturbed feature to the perturbed feature. Such a map is referred to as a *continuation* of a specific element into the perturbed system.

1.2.2 KAM theory

In essence, KAM theory is a perturbation theory, as it concerns the continuation of quasi-periodic invariant tori. It follows the same basic principles and is structured the same as general perturbation theorems described above. As such, some of its applications are largely analogous in form and function to general perturbation theories. However, the construction and conditions used by KAM theory are quite different from the methods used by more general perturbation theories.

To start, the context of KAM must satisfy certain conditions, integrability within KAM theory is its own concept somewhat different to the more classical definitions. Additionally, smoothness of the conjugation in certain variables is determined in the Whitney sense, instead of more familiar definition based upon open subsets of the domain. This must be done in this way, because the domain in these variables is closed and does not contain any open subsets.

The result is peculiar in its own way, the persistence of single Diophantine tori under any small perturbation is not assured. As it is possible to construct arbitrarily small perturbations that shift any frequency to a resonant one. Therefore most KAM theorems prove that of a family of quasi-periodic invariant tori, satisfying a non-degeneracy and frequency condition, a "large" subset of tori survives under small perturbations. The definition of the family of tori sometimes requires additional external parameters, which also need to be considered.

Here, the term "large" means large in the measure-theoretic sense. The measure of the subset of surviving invariant tori is non-trivial[9] in the set of all considered invariant tori. More specifically, it is known that for $\tau > n - 1$ the set of Diophantine invariant tori, the invariant tori for which the persistence is proven by KAM theory, is positive in the set of all considered invariant tori. This is because the Lebesgue measure of the subset of Diophantine frequency vectors within $\mathbb{R}^n$ is positive.

In addition to the examples given above, there are quite a few additional considerations and subtleties to KAM theory.

[9] Or equivalently positive.

1.2.3 Applications of KAM theory analogous to perturbation theory

Perturbation theorems enable the deduction of many properties of the perturbed system by analyzing the properties of the simpler system. This, of course, has many applications, and it is not surprising that perturbation theory is one of the most studied theories in dynamical systems

Simplification of complex dynamics

As stated in section 1.2.1, in most perturbation theorems, the class of integrable systems is defined to have well-understood dynamics, from which clear conclusions can be drawn. A perturbation theory aims to prove that these well-understood dynamics are preserved under certain perturbations.

Some difficult and subtle complex dynamics can be reinterpreted as a perturbation of an integrable system. If this perturbation is sufficiently small[10], the perturbation theory implies that the dynamics are similar to the integrable system. In this way, much knowledge about the complex dynamics can be obtained from the well-understood integrable system.

Astronomy

The classical example of this application and the original motivation for the first perturbation theorems is the study of the solar system. Consider the classical 3-body problem with the Sun, Mercury and Jupiter[11]. The complete system is too difficult to be exactly solvable, however the interaction between Mercury and Jupiter is relatively weak when compared to the Sun-Mercury and Sun-Jupiter interactions.

When calculating the orbit of Mercury over time, the Mercury-Jupiter interaction can be interpreted as a perturbation of the easily solvable system, which only includes the Sun-Mercury and Sun-Jupiter interactions. The solution of the orbit of Mercury can be further refined by including the more well-understood and easily calculable parts[12] of the perturbation in the approximate simplified system.

Classical perturbation theory made it possible to make successively more accurate predictions on the orbits of our solar system. Despite this success, classical perturbation theory was never able to decisively prove long-term stability.

A famous example and one of the first applications of KAM theory by Arnol'd was the proof of long-term stability of specific cases of the 3-body problem. We quote Arnol'd [2];

> If the masses, eccentricities and inclinations of the planets are sufficiently small, then for the majority of initial conditions the true motion is conditionally periodic and differs little from Lagrangian motion with suitable initial conditions throughout an infinite interval of time $-\infty < t < \infty$

Sufficiently small is not to be taken lightly, explicit estimates of the thresholds were studied by Hénon [44]. When applying classical KAM theory[13] to the 3-body problem, the threshold of mass–ratio of the primaries is less than 10^{-333}, and using Moser's theorem less than 10^{-48}. For comparison, the mass-ration between Jupiter and the Sun is around 10^{-3}.

Other fields

Astronomy is not the only field where perturbation theory can be applied. Within quantum theory, there are complex dynamical systems that model many particles and their interactions. An example of such systems is the model for large molecules, such as proteins. Some of the interactions between far-away particles can be considered a perturbation, similar to far-away planets in orbital mechanics. The application of perturbation theorems in this way is common in many different fields.

KAM theory can also be applied in a similar way. It conjugates the quasi-periodic tori of an integrable system to the quasi-periodic tori within the unperturbed system, thus ensuring the existence of such tori in the perturbed system. These perturbed tori then have well-understood long-term dynamics and provide important insights into the long-term behavior of the entire system.

However, the tori that KAM theory conjugates are specific and probabilistic in nature. Classical KAM theory also has notoriously small thresholds on the size of the perturbation, making it difficult to apply to such situations. Moreover, the dynamical systems studied by KAM theory are required to be similar to the integrable systems KAM considers, i.e. having quasi-periodic tori in some form. Examples include very specific quantum field theories and complex quasi-periodic celestial mechanics. This makes KAM theory quite a niche tool within physics.

[10] The maximum (or equivalently infinity) norm of the perturbation is bounded by a very small value.
[11] Chosen as example for the relatively small interaction between Mercury and Jupiter.
[12] Such as linearization or other truncations of the perturbation.
[13] Arnol'd's original KAM theorem, as stated in his 1963 study [1].

Bifurcation and normal form theory

A more abstract application of perturbation theory involves the study, detection, and persistence of bifurcations. This is typically achieved through the construction of normal forms for bifurcation scenarios. Almost all bifurcating elements in a system can be locally conjugated to a specific normal form that exhibits the same bifurcation. This normal form is typically chosen to have desirable symmetries and properties that simplify the study of the bifurcation.

Normal form theory.

A well-known example of a normal form theory is the Hartman-Grobman theorem, which locally conjugates the original system around a fixed point with the linearization around this point. This simplifies the study of a great many equilibria to the study of the normal form. Degenerate equilibria should be normalized to higher-order normal forms, as in this case, the linear terms after reduction are zero. In the Hamiltonian context, Birkhoff normal form theory is the most well-known example.

Normal forms can still be quite complex, with many higher-order terms. Perturbation theory allows these higher-order terms to be neglected by treating them as perturbations of the truncated normal form.

Quasi-periodic bifurcation theory

Quasi-periodic invariant tori, as dynamical features analogous to fixed points and periodic orbits, have a largely similar bifurcation theory associated with them. This is known as quasi-periodic bifurcation theory, and the application of KAM theory is a fundamental part of it.

1.2.4 Applications specific to KAM theory

KAM theory extends beyond perturbation theory and has applications and implications in mathematics that are not directly comparable to general perturbation theory.

In particular, KAM theory yields significant results in the study of long-term behavior of dynamical systems, as well as the probabilistic persistence of quasi-periodic invariant tori. These results have significant implications for ergodic theory and statistical mechanics. All of these factors contribute to the significance and importance of KAM theory.

Ergodic theory

Perhaps the most significant impact of KAM theory is its implication to ergodic theory. Ergodic theory studies the long-term distribution of orbits and is the foremost method for studying chaotic behaviors, and has been very influential in thermodynamics and statistical mechanics.

Before the advent of KAM theory, it was often assumed that all nearly integrable Hamiltonian systems with enough degrees of freedom are ergodic on fixed energy levels, implying that the long-term distribution of orbits is uniform on those levels. However, KAM theory contradicts this hypothesis for nearly integrable Hamiltonian systems with quasi-periodic tori. Orbits get trapped on these tori and become unable to visit any points outside of them, thereby disproving ergodicity.

As a result, a large number of systems have been excluded from being assumed as ergodic due to the persistence of invariant tori.

1.3 Difficulties with common persistence theory techniques

Consider the following perturbed analytic vector field $X_\varepsilon = X + P_\varepsilon$ on the manifold M, where P_ε is an analytical perturbation bounded by ε on a neighbourhood D containing the feature or element, i.e. $\sup_{x\in D} ||P_\varepsilon(x)|| = ||P_\varepsilon||_D < \varepsilon$ where $||\cdot||$ is the maximum norm defined by $||x|| = \max_{0\leq i\leq n} |x_i|$ for every $x \in \mathbb{R}^n$. The flow of X gives rise to the system

$$\dot{x} = X_\varepsilon(x). \tag{1.2}$$

1.3.1 Implicit function theorem

One of the most commonly used tools to ensure the continuation of elements is the implicit function theorem, which is directly applicable to many simple elements such as fixed points and periodic orbits.

Furthermore, as the conditions on the system are quite mild, the theorem has been successfully applied to prove the persistence of elements in a wide range of contexts. Due to its widespread applicability, the implicit function theorem remains an essential tool in the study of dynamical systems today. Please refer to the appendix B for a full statement of the implicit function theorem.

The classical analytic implicit function theorem relies on the use of Taylor series of X_ε in the perturbation variable ε about $\varepsilon = 0$, given by

$$X_\varepsilon = X + \varepsilon P_1 + \varepsilon^2 P_2 + O(\varepsilon^3). \tag{1.3}$$

For sufficiently small ε, the higher-order terms become negligible. Thus, given $X(x_0) = 0$ for $x_0 \in D$, we can inductively construct a function $\sigma : \varepsilon \mapsto x$ such that $X_\varepsilon(\sigma(\varepsilon)) = 0$ and $\sigma(0) = x_0$.

This approach is particularly convenient for fixed points, as they are identified by $X(x_0) = 0$. It is easy to see that they are continued along the path of $\sigma(\varepsilon)$ as ε varies. The theorem can also be easily applied to Poincaré maps of periodic orbits, where the fixed points correspond to the periodic orbits.

The breakdown of the implicit function theorem

The classical implicit function theorem is not applicable to quasi-periodic orbits. The Poincaré maps of such orbits result in quasi-periodic circle diffeomorphisms, which do not have any fixed points or invariant sets that can be mapped to a single point.

Let us consider the simplest case, the 2-torus. Suppose we take the Poincaré map, which is a quasi-periodic circle map. Under the repeated application of this map, the resulting sequence would never visit the same point twice. Even worse, the sequence would come arbitrarily close to every point in the Poincaré section eventually. Therefore, there are no fixed points and there are no invariant proper subsets that could be contracted to a single point.

Without any isolated fixed points in the Poincaré map, no Taylor series can be constructed around such hypothetical points. As the classical implicit function theorem relies on such a Taylor series, they cannot be used for quasi-periodic tori.

In recent years, generalized implicit function theorems for quasi-periodic tori have been developed. However, all of these rely on KAM theory methods for their proof.

1.3.2 Invariant manifold theorems

Persistence of invariant subsets does not always require strict orbital equivalence. It can also be valuable to determine whether general invariant sets survive, without considering their internal dynamics. These sets could be points, orbits, tori, and other general invariant subsets of the phase space.

One major theorem [32] by Fenichel, known as the normally hyperbolic invariant manifold theorem, proves the persistence of normally hyperbolic invariant manifolds within general systems. This theorem is widely applied and has been successful in answering many questions, including determining the persistence of invariant tori under small perturbations. For further information on invariant manifolds and their persistence, consult the book [50] by Hirsch, Pugh, and Shub.

However, it is not capable of determining the persistence of quasi-periodic invariant tori. This is because center manifold theory does not determine whether the parallel flow of the perturbed torus remains quasi-periodic. The perturbed flow, while still being close to the original parallel flow, is not guaranteed to be quasi-periodic.

Additionally, under general perturbations, it is possible to change the frequency of an invariant torus, and it might change from quasi-periodic to resonant or to an even more unpredictable non-linear flow.

Further restricting the size of the perturbation is not enough. Since, analogously to the rational numbers being dense within the real numbers, resonant frequencies are dense within all frequencies. This means that for every quasi-periodic frequency, there is a resonant frequency arbitrarily close. So no matter how small the bound is,

there always is a perturbation within that bound that can change the frequency from non-resonant to resonant. This risk of losing quasi-periodicity with any small perturbation makes it difficult to prove the persistence of quasi-periodic tori.

1.3.3 Conjugation

A common approach used to prove the persistence of a dynamical feature or element is to construct a conjugation between the integrable and perturbed system.

Let X be a vector field on M that induces dynamics with a dynamical feature of interest, which is referred to as the element, and let $X_\varepsilon = X + \varepsilon P$ be a small perturbation of X.

Definition 1.3.1 (Local conjugation)**.** Let $\phi : M \to M$ be a homeomorphism and $U \subset M$ be an open neighbourhood of the element. Suppose that the diagram

$$\begin{array}{ccc} U & \xrightarrow{\phi} & \phi(U) \\ {\scriptstyle \varphi_X^t}\downarrow & & \downarrow{\scriptstyle \varphi_{X_\varepsilon}^t} \\ U & \xrightarrow{\phi} & \phi(U) \end{array}$$

commutes where φ_X^t is the flow of X, and $\varphi_{X_\varepsilon}^t$ the flow of the perturbed vector field X_ε. Then it is said that ϕ *locally conjugates* the original and perturbed system and the systems are *locally topologically conjugate*. Additionally, if M is equipped with a structure[14], ϕ must preserve said structure.

The local conjugation implies that X and X_ε are topologically equivalent[15] on U and $\phi(U)$. This equivalence implies the persistence of every element and orbit within U, instead of just the persistence of a single element. Care needs to be taken to ensure that the (possibly) persistent element is contained within U, as the systems may differ significantly outside the narrow bounds of U and $\phi(U)$. The construction of such conjugations can be very challenging, depending on the persistent element and context.

Difficulties with conjugation (small denominators)

Proving the persistence of quasi-periodic orbits cannot be accomplished using the classical implicit function theorem or the center manifold theorem. Therefore, we must take a more robust approach and prove persistence by conjugation.

The general approach to constructing such a conjugation is to approach the conjugation inductively, where at each iteration, we solve a linearization of the conjugation relation to obtain an approximate conjugation. Each of these approximate conjugations moves the vector field X_ε closer to X, and by taking the limit of these approximate conjugations, we obtain ϕ.

This method is also used in the proof of KAM theory, but additional complications arise. Solving the linearization of the conjugation relation is done by means of the Fourier series, which results in a formal Fourier series for the approximate conjugation.

For general frequencies, this formal series is not guaranteed to converge to a well-defined approximate conjugation. This is because small denominators occur within the coefficients of its higher-order terms of the Fourier series. These denominators can quickly become smaller as the order increases, causing the coefficients to blow up faster than the natural decay of higher-orders can decrease the term. Therefore, the higher-order terms do not necessarily approach zero, and the convergence of the entire series is in doubt.

Truncating the Fourier series to approximate the approximate conjugation does not result in a usable approximation, as the truncated higher-order terms can significantly impact the actual conjugation, making the approximation very inaccurate.

This small denominator problem has remained difficult to deal with for quite some time, and a breakthrough was needed to solve this problem.

[14] Such as the sympletic structure in case of Hamiltonian dynamics.

[15] Other stricter notions of equivalence between systems, such as *smooth* or *orbital* equivalence, do exist. However, they are not necessary for studying the persistence of elements.

1.4 Development of KAM theory

As with many influential theories, KAM theory and small denominator problems in general have a long history. This section is dedicated to exploring the origins of small denominators and KAM theory.

1.4.1 Historical background

KAM theory has its origins, like many problems in dynamical systems, in astronomy, particularly in the study of the solar system and predictions of the orbits of the moon and Earth. These were of great interest to astronomers and mathematicians of the 17th and 18th centuries due to their importance in the astronomical navigation of maritime vessels. It is not surprising that this is where the first Hamiltonian perturbation theories were developed.

Classical

The most well-known of these scholars that studied the 3-body problem was Newton, who was the first to derive Kepler's laws. Newton was able to find a way to approximate solutions of this system, by means of the brand-new perturbation theory. However, long-term behavior, and particularly the stability, of the 3-body problem remained unsolved for a very long time. Ever since the long-term behavior has been one of the most premier problems. The work of Laplace and Lagrange on the problem, including the discovering of Lagrange points within the solar system, tentatively indicated that stability could eventually be proven.

Throughout the nineteenth century, astronomers, physicists, and mathematicians continued to use perturbative methods extensively, systematically including all interactions within the system as further perturbations. Early methods were refined, tested, and further developed into a trigonometric series expansion for solutions of the 3-body system. This remarkable series, known as the Lindstedt series, was introduced by Lindstedt, Newcomb, and Poincaré, and remained the most effective approximation for some time. The method was popular and eventually pushed to its limits by the astronomers of the time.

The most impressive example of this is the work of C. E. Delaunay. In his studies of the 3-body problem, which were published in two 900 page volumes in 1860 and 1867, he used perturbative methods to calculate the orbit of the moon under influence of the sun. After calculating 505 perturbation terms, he was the first to notice that higher order terms can be greater than the first order terms. In his investigation of this phenomenon, Delaunay gave the first description of small denominators. It was apparent that these small denominators were a problem, as he noted that the large higher-order terms rendered the truncated Hamiltonians useless in their predictive power. The convergence of the Lindstedt series remained unproven in the 19th century and was one of the major questions of the time.

Poincaré and Siegel

On the occasion of his 60th birthday, King Oscar II of Sweden established a substantial prize to anyone who could definitively prove the convergence of the Lindstedt series, or if the problem was unsolvable, to anyone who could make a significant contribution to the field. If convergence had been proven, it would have meant that the n-body problem could be solved analytically.

Poincaré took on the King's challenge and eventually won the prize. Not for solving the problem but for seemingly demonstrating that the Lindstedt series is divergent. He showed that the series could not converge to a periodic solution for arbitrary constants.

Poincaré eventually became aware of the small denominators that plagued the Lindstedt and similar other series. He frequently made incorrect conjectures involving small denominators, one of which was about the divergence of perturbation series about invariant tori with fixed frequencies, which, in fact, was proven to be false by KAM theory. However, Poincaré never claimed to outright disprove convergence[16].

Again, the convergence of the Lindstedt and similar series remained unsolved for quite some time. Eventually, the next step was made by Siegel. In 1942, he published his study "Iteration of analytic functions," where he decisively proved the convergence of formal linearizations of specific complex analytic functions.

In the study, he introduced the Diophantine conditions to control the size of the small denominators in the formal linearization, which otherwise could cause the series to diverge. This was the first result that proved convergence of a series that still included small denominators. Before Siegel, it was simply not known that small denominators could be controlled, and most formal solutions containing them were just assumed to diverge or, at least, that their convergence could not be proven.

This was a major innovation in dealing with small denominators. This problem has remained relevant well after Siegel's study and is covered in more detail in section 2.1.3.

[16] See the provided quote in Herman's survey [45] on the complex linearization problem, which was omitted here for brevity.

1.4.2 Kolmogorov

The breakthrough in the persistence of quasi-periodic tori in Hamiltonian systems came in 1954 when Kolmogorov gave a lecture in Russian at the International Mathematics Conference on his paper "On the conservation of conditionally periodic motions for a small change in Hamilton's function."[17] In the lecture, Kolmogorov discussed quasi-periodic motions confined to invariant Lagrangian tori within Hamiltonian systems. The existence of such quasi-periodic tori was already well-known in Liouville-integrable systems. Kolmogorov examined their persistence under small non-integrable perturbations.

This is commonly seen as the start of KAM theory and monumentally the first instance of the KAM method for dealing with small denominator problems.

Diophantine conditions for vector fields

Kolmogorov was the first to construct[18] a conjugation between the tori in the integrable and perturbed Hamiltonian systems. To achieve this, Kolmogorov introduced several conditions on the integrable system and the quasi-periodic invariant torus contained within it.

The most important of these were the Diophantine conditions on the frequencies. Inspired by Siegel's work, Kolmogorov used a vector field variation of the Diophantine conditions to circumvent the small denominators within the formal solution of the conjugation. For a complete description and properties, see section 2.2. The second condition was a non-degeneracy condition that was required for the proof. More on this is explained in section 3.5.

KAM method

The second and perhaps more important innovation was to use an adapted Newton's algorithm[19] to construct the conjugation. Each iteration of the algorithm produces a better approximation of the desired conjugation.

The construction of these approximations still contained small denominators, which had to be managed by Diophantine conditions. Moreover, successive iterations could perturb the frequency when the approximate conjugations were not appropriately chosen, thus complicating the use of the Diophantine conditions.

Kolmogorov solved these problems by keeping the frequency fixed for each iteration step. This was achieved by shifting the action variable with each approximation, such that the frequency remained fixed. The fixed frequency allowed for the systematic use of the Diophantine conditions, as only one frequency had to satisfy the conditions. Without this correction, each iteration would result in a slightly perturbed frequency that then had to be verified to be Diophantine. This method of constructing the approximation is now known as Kolmogorov's scheme.

The additional parameter shift did introduce a small additional error to the approximate conjugation at each step. Fortunately, the quadratic convergence of the KAM method was strong enough to compensate for slightly less optimal approximations.

Issues

The paper, unfortunately, did not come without issues, mostly due to its brevity. Kolmogorov's paper is only four pages long, which is quite typical of the Russian tradition of mathematics. However, despite containing the prototype of KAM theory, it led to some controversy. As most later KAM proofs can easily number into 30 pages, the four pages of Kolmogorov were a bit sparse on details.

For example, consider the generating function required to construct each transformation at every step of the iteration. The convergence of the limit of each canonical transformation is largely glossed over in Kolmogorov's paper, only noting that it was "subtle" in nature. Kolmogorov also omitted most of the estimates required to formally prove the convergence of each step.

It took further fundamental work to convince the world of its correctness. Therefore, most modern interpretations of Kolmogorov's method include much more than what was actually written down. For a statement of Kolmogorov's theorem and an interpretation of the outline set in Kolmogorov's paper, see the studies by Chierchia [25, 26].

[17] Translated from Russian.
[18] Or, depending on the standard of rigour, outline the construction of such a conjugation.
[19] This method is also referred to as the quadratic iterative method or more commonly as the KAM method.

1.4.3 Arnol'd

Another significant milestone in the development of KAM theory occurred almost ten years after Kolmogorov's initial work, on the occasion of his 60th birthday. In 1963, Arnol'd published the first comprehensive proof [1] of the theory.

Improvements to rigour

One of the major steps that Arnol'd made in his study was filling in the gaps left in Kolmogorov's proof. Arnol'd explicitly proved many of the estimates required for convergence, proved the convergence of the generating function, and made many improvements to the rigor of the proof.

Different iteration scheme

Arnol'd did not just improve on the results of Kolmogorov, he also innovated. He used a slightly different iteration scheme, which was still based on the adapted Newton method. This scheme is more classical in nature, which is somewhat easier algebraically but more delicate in terms of the required estimates.

The construction of the approximations and the application of the Diophantine conditions is notably different in Arnol'd's scheme. Arnol'd did not fix the frequency to a single Diophantine value; instead, at each iteration step, the frequencies are restricted to a domain on which they behave well for the calculations at hand. The calculation for the construction of the approximation is simplified by truncating the Fourier series at a set cut-off[20] for each iteration. The simplified calculation only deals with a finite number of small denominators at each iteration step, thus limiting the conditions required on the frequencies and allowing for larger domains. This also allows for some limited frequency shift at each iteration, as long as the shifted frequencies are still contained within the chosen domain. The domains of the frequencies eventually approach the set of Diophantine frequencies, thus leading to the same Diophantine conditions as Kolmogorov's scheme.

These ideas and the scheme in general have since been widely adopted in KAM theory, and notably in infinite-dimensional Hamiltonian perturbation theory.

Arnol'd originally stated his KAM theory on the persistence of cantorized sets of quasi-periodic orbits, as the scheme allows for the conjugation of the entire frequency domain at once. This opened the door for investigating regularity in the frequency parameter, as well as other re-parameterizations of the frequency.

New non-degeneracy condition

As an addendum to his proof of KAM theory, Arnol'd also introduced a novel non-degeneracy condition. This non-degeneracy condition reduces the required parameters in KAM theory by one, and it is the first of many parameter-reducing optimizations in KAM theory. Parameter reduction has since become one of the major research topics within the KAM literature.

1.4.4 Moser

The last mathematician immortalized in the name of KAM theory is J. Moser. Moser has made multiple important contributions to the KAM literature; however, the most well-known of his studies is his paper [59] on a differentiable case of KAM theory, which gave the M to KAM theory.

Differentiable case

In 1962, Moser published a KAM theorem regarding the area–preserving twist mappings of an annulus [59]. This particular study was remarkable as the perturbation was not required to be analytical, but could be finitely differentiable instead. Initially Moser chose C^{333} as minimum differentiability, but this was later brought down to just C^5 by Rüssmann in the two dimensional complex case. Finitely differentiable KAM theory has since been generalized to n dimensions.

The additional difficulty with the finite differentiability of the perturbation is that, when using the regular iteration method by Kolmogorov and Arnol'd, at each iteration step the approximated conjugation yields a new perturbation which has one less degree of differentiability. If left unchecked, this would eventually introduce discontinuities in the conjugation and subsequently break the KAM method.

Moser resolved this issue by applying a smoothing technique inspired by Nash, which uses convolutions to restore a certain degree of differentiability at each step of the iteration. This ensured each new perturbation remained suitably differentiable, which solved the loss-of-continuity problem.

The extra smoothing step within each iteration introduces a small additional error to the approximations, but again the convergence of KAM method is strong enough to compensate for this.

[20] This cut-off is today referred to as the *ultra-violet* cut-off.

1.5 More recent contributions

It is important to note that KAM theory has undergone many generalizations and refinements beyond the ones mentioned in this thesis. In fact, there are numerous results in the literature that this thesis[21] cannot cover exhaustively. However, a brief list of some of the more influential and commonly used results is included here.

Generalization to other conservative systems

The original KAM theory was developed in the Hamiltonian context, but it was quickly generalized to three other main contexts: dissipative, volume-preserving, and reversible. These generalizations are now well-known, and most recently published KAM theories can be readily extended to each context. Section 3.1.1 covers these contexts in more detail, but for further information, the book by Broer, Huitema, and Sevryuk [17] provides a comprehensive discussion of the contexts and the associated KAM theorems.

Lie algebra approach

It is well-known that Lie algebras and Lie groups can be used to represent structure-preserving vector fields and diffeomorphisms. This interpretation provides a unified approach to KAM theory for multiple contexts and can lead to some simplifications, though not many. Moser was the first to discuss this in his studies [60, 61], as part of his investigation into KAM theory for various contexts.

Later on, Moser and his student Pöschel perfected the KAM proof. This proof introduced the now familiar notation for the linearized conjugation equation[22]. Solving this equation results in a linearized conjugation, which additionally preserves the underlying structure of the context. Combining many linearized conjugations eventually results in the full conjugation constructed in KAM theory. In the Hamiltonian context, this method is conceptually simpler[23] to work with than the more traditional generating functions method used to construct structure-preserving approximate conjugations.

Today, the Moser-Pöschel proof serves as a framework for proving general KAM theorems, and many KAM theory proofs are adaptations of this proof.

Whitney regularity

Whitney-differentiability is a refinement of the classical notion of differentiability, which is only defined for functions on open sets. Whitney-differentiability extends this concept to functions on arbitrary, possibly non-smooth, closed sets, such as the set of Diophantine frequencies. Whitney-differentiability was introduced to KAM theory by Moser, Pöschel, and Zehnder in the 1980s[24], and has since become a fundamental tool in the field.

Parameterized KAM and parameter reduction

Modern KAM theory is often formulated in the parameterized variant, where each invariant quasi-periodic torus is parameterized by internal and external parameters. Parameterized KAM theory is easier to prove and is more widely applicable; some contexts even require the presence of external parameters to state meaningful KAM theorems. This approach to proving KAM theory was first discussed by Moser [61] in 1967. See section 3.5 for more information.

Parameterized KAM theorems with a minimal number of external parameters are more useful in many situations. Therefore, the development of parameter reduction techniques has been a major research focus in KAM theory, and many different methods have been proposed over the years. These methods involve the weakening of the non-degeneracy conditions. For multiple example of weakened non-degeneracy conditions see the surveys [39, 81] by Hanßmann and Rüssmann.

Optimal result

An optimal result for small denominators was obtained for the complex linearization problem by Yoccoz [100]. Although, this topic is not KAM theory, it still is closely related to KAM theory. For more information on this topic, please refer to section 2.2.

[21] Or any other surveys and tutorials in KAM theory to the best knowledge of the author.
[22] Also commonly referred to as the homological equation.
[23] While this is the general consensus, personal preferences can vary.
[24] The exact first introduction of Whitney-differentiability to KAM theory is unknown. However, the earliest published application that the author is aware of is in the 1982 study by Pöschel [71].

Lower dimensional tori

Closely associated with the generalization to different contexts is the study of the persistence of lower-dimensional quasi-periodic invariant tori. These tori have less dimensions than the base space and can be seen as an embedding. This implies that these tori have normal bundle, on which non-trivial normal dynamics are common. The normal dynamics of an invariant torus are studied in very small neighborhoods of the lower-dimensional invariant torus and are often linearized around this torus. The normal dynamics add another layer of complexity to KAM theory.

Most notably, the internal and normal dynamics of quasi-periodic invariant tori are not independent from each other and interact in non-trivial ways. The study of these interactions was pioneered by Mel'nikov [58], who first stated the normal-internal non-resonance conditions required for persistence.

This direction was further developed by Braaksma, Broer, Huitema, and Takens, among others. Their work culminated in the comprehensive studies on the unfoldings of quasi-periodic tori contexts, such as [10, 18]. This theory eventually found its way into books on quasi-periodicity and chaotic dynamical systems, such as [17] and [20], respectively.

A survey by Broer [11] summarizes his and his colleagues' work on the KAM theory of lower-dimensional quasi-periodic tori. This survey presents a comprehensive view, making it a very useful reference for researchers and students interested in KAM theory.

Adjacent research

KAM theory and its methods have also been adapted and applied in many research topics that should be seen as separate fields with their own extensive literature. As this thesis is confined to KAM theory in the classical sense, most adjacent research has been excluded. However, I do name three adjacent fields, which are three of the most closely related to KAM theory in the classical sense.

KAM theory for infinite-dimensional Hamiltonian systems, which is also referred to as weak KAM theory. The techniques used to prove weak KAM theorems are unsurprisingly very similar to KAM theory methods. Weak KAM theory remains an active research area. For more information, see [30].

Nekhoroshev theory provides an estimate of the time for which nearly integrable Hamiltonian systems remain stable, i.e. the time it takes for an orbit near the invariant torus to leave a small neighborhood of said torus. The theory is named after the Russian mathematician Nikolai Nekhoroshev, who first introduced it in the early 1970s, and formally published it in 1977 [63]. This theory finds widespread applications in physics, as the bound on the size of the perturbation for which meaningful results hold is much greater than the typical bound for KAM theory. The proof of the theorem is also remarkably similar to that of general KAM theorems. For more information see [36].

Lastly, we consider quasi-periodic bifurcation theory, which studies bifurcations of quasi-periodic invariant tori. Since the small bifurcation parameter permutes the system at the bifurcation point, KAM theory is applicable. By applying KAM theory, it is known that quasi-periodic invariant tori only persist for certain small bifurcation parameters, resulting in Cantorized bifurcation diagrams. Due to the more complex normal forms of these bifurcations, quasi-periodic bifurcation theory requires its own variants of KAM theory. For more information see [38, 40].

The tutorial [28] gives a quite extensive list of research topics in KAM and adjacent literature from 1992 to the time of publishing (2001) in the introduction.

Further reading and tutorials

Dumas in his book [29] introduces KAM theory history, applications and significance in much greater detail and is more accessible and friendly to new students than most of the references in this thesis.

The KAM tutorial [28] by De la Llave, the lecture notes [73] by Pöschel, the introduction to small denominators [56] by Marmi and finally the introductory text [19] by Broer and Sevryuk form the basis of the study of KAM theory. It is recommended that anyone unfamiliar with KAM theory consult these introductory texts, as they provide a good exposition on the KAM method and small denominators.

The books [17, 20, 38] by Broer, Hanßmann, Huitema, Sevruyk and Takens provide more context and application for KAM theory, and [17, 38] include a proof of a KAM theorem in the appendix.

Lastly, for the original proof and context of Kolmogorov's, Arnol'd's and Moser studies, consult the historic examination of the Kolmogorov's, Arnol'd's and Moser's proof [26] by Chierchia. Chierchia also published in [25] Kolmogorv's orginal proof with all estimates made .

As mentioned before, this introduction to KAM theory is by no means exhaustive, and further reading is necessary to fully comprehend the theory.

Part I: Small denominator problems

This section is dedicated to small denominator problems. These are also known as small "divisor" problems, but this is the result of an old mistranslation. The small "divisor" problem is not concerned with the divisibility of terms, as the term "divisor" would imply, but with the presence of small denominators in formal series. This thesis refers to the problem as the small denominator problem.

Small denominators occur within KAM theory and in a wide array of other problems concerning quasi-periodicity. These problems can be roughly classified by their context, by small denominator problems for maps and small denominator problems for the flow of vector fields.

The small denominator problem for maps is quite different in notation and conditions from the one for the flow of vector fields, but the principles that govern the solution are more or less the same. An example of the problem where small denominators for maps occur, is the complex linearization problem, which is discussed in more detail in section 2.1.

Let us consider a common problem within KAM theory, known as the one-bite small denominator problem, which is a very common form of small denominator problems in the flow of vector fields context. Let $f : \mathbb{R}^n \to \mathbb{R}$ be some bounded function, $\omega \in \mathbb{R}^n$, and consider

$$D_\omega u = \left\langle \omega, \frac{\partial}{\partial x} u \right\rangle = f, \tag{2.1}$$

where we solve for $u : \mathbb{R}^n \to \mathbb{R}^n$. Directly solving the equation results in

$$u(x, \omega) = \sum_{k \in \mathbb{Z}^n \setminus \{0\}} f_k(\omega) \frac{e^{i\langle x,k\rangle}}{i\langle \omega, k\rangle}. \tag{2.2}$$

We are concerned for which frequency vectors $\omega \in \mathbb{R}^n$ this series converges, as the convergence of (2.2) implies that the function u is well-defined.

It is already apparent that the resonant frequencies are a problem in this bare-bones problem, as these result in $\langle \omega, k\rangle = 0$ for some $k \in \mathbb{Z}^n$. Division by zero is not well-defined and these frequencies are to be excluded for the series to be well-defined.

However, even with all resonances excluded, it is possible for all $k \in \mathbb{Z}^n$ to find an ω in the remaining frequencies such that $\langle \omega, k\rangle$ is arbitrarily small. Therefore it is possible that for a randomly chosen non-resonant ω, there are very small denominators $\langle \omega, k\rangle$ present at more or less random values of k in the sum. The greatly amplified terms could cause the sum to diverge or at least complicate the question of convergence greatly.

The terms $i\langle \omega, k\rangle$ are the so-called small denominators in this version of the problem.

Outline of part I

In the first section, we explore the first small denominator problem that was adequately solved: the linearization problem. We briefly cover the early solutions to the linearization problem and the occurrence of a small denominator problem for maps.

Afterward, we cover the Diophantine conditions and discuss their properties. We also provide a more recent generalization in the form of approximation functions.

Section 2.3 provides an introduction to Whitney differentiability and the associated extension theorem. Although these concepts are not directly required to solve small denominator problems, they are nevertheless a common feature of modern KAM theory.

Lastly, we fully cover the Whitney-smooth one-bite small denominator problem. The one-bite problem is arguably the simplest small denominator problem, having a very "linear" character and being directly solvable by the Fourier series. However, here it is combined with the use of the inverse approximation lemma to give its Whitney-smooth formulation. This method is analogous to the methods used in complete KAM proofs to prove regularity.

2.1 Complex linearization problem

The first successful solution to a problem containing small denominators was Siegel's theorem for the complex linearization problem in one dimension, which is sometimes referred to as the Schröder-Siegel problem. This section explores this problem and a few landmark results associated with it.

Although this problem is not vital to understanding KAM theory itself, as the problem does not necessarily concern itself with specific dynamical settings, regularity and parameters, and was even originally solved with methods very different from the now common KAM theory methods.

Nevertheless, it is still closely related to KAM theory as the same KAM method developed by Kolmogorov can be used to deal with the small denominators. This reduces the KAM method to its most basic form, where only the minimal amount of geometric considerations have to be taken into account. Vice versa, the Cauchy majorant method, or direct method, has also been applied to KAM theory itself by Rüssmann [76] and Zehnder [104]. There has been significant cross pollination of methods and concepts between the two problems, and therefore it is useful to be aware of the linearization problem when studying KAM.

Despite the similarities between the linearization problem and KAM theory, they are not equivalent. The setting of the linearization problem concerns the linearization of a fixed point, while KAM theory concerns the persistence of quasi-periodic tori. This difference in setting also requires different Diophantine and non-degeneracy conditions.

As mentioned before, the linearization problem, or more accurately the linearization of holomorphic diffeomorphism in 1 dimension with an indifferent fixed point, was the first small denominator problem to be properly solved. This is commonly marked as the start of KAM theory, and was paradigmatic for both KAM theory and holomorphic dynamics. Since then, several other developments have been made on the subject, such as the first optimal results for Diophantine frequency conditions and the development of novel methods. Several of these have since made their way to KAM theory.

2.1.1 Problem of linearizability

We start with the setting of the problem and the occurrence of small denominators within it. Consider the following situation.

Definition 2.1.1. Let $f \in (\mathbb{C}[[z_1, z_2, \cdots, z_n]])^n$ be a germ of a diffeomorphism of $(\mathbb{C}^n, 0)$. Denote $z = (z_1, z_2, \cdots, z_n)$, then f can be written as

$$f = Az + O(z^2), \tag{2.3}$$

with $A \in \mathrm{GL}(n, \mathbb{C})$. Denote the eigenvalues of A by $\lambda_1, \cdots, \lambda_n$, where eigenvalues could be equal.

We give a definition of linearizability.

Definition 2.1.2 (Linearization)**.** Let f be as defined in definition 2.1.1. Then if there exists a unique (formal) diffeomorphism h such that, $h = z + O(z^2)$ and the following equation[1] is (formally) satisfied

$$h^{-1} \circ f \circ h(z) = Az \tag{2.4}$$

then f is (formally) linearizable in a neighbourhood of 0.

This allows us to state the complex linearization problem in n-dimensions.

Problem 2.1.3. Suppose f is a germ of a holomorphic n-dimensional diffeomorphism at 0, with $f(z) = 0$ and $A = Df(0)$. Can f be linearized in a neighbourhood of 0?

This problem is important, as the linearizability is closely related to the stability of a point. In the linear dynamics, it is relatively easy to determine the stability of a point. Since the linearized system is by definition conjugated to the original system, they are locally topologically equivalent[2] and the stability result can be applied to the original system.

Therefore, it is not surprising that linearization has become an important research topic, and considerable efforts have been made to classify holomorphisms based on their linearization properties.

The problem has been studied for quite some time, the first description of the problem in 1 dimension given by Schröder [85] in 1871. The problem in 1 dimension has been referred to as the Poincare-Schröder or the Schröder-Siegel problem[3].

[1] This equation is referred to as a homological equation by some authors.
[2] For further reference, see [35].
[3] Named after the first description and the first solution, respectively.

2.1.2 Solution to the linearization problem

The linearization problem has been a major area of interest in the field of dynamics, with significant effort directed towards its solution. Prior to the development of Diophantine conditions and small denominators, only specific cases of the problem had been successfully solved.

Eigenvalue conditions

These specific cases were defined by conditions on the eigenvalues of A. Three examples are given. Let λ_j be the eigenvalues of A, and consider the following condition:

$$\sup_j |\lambda_j| < 1 \quad \text{or} \quad \sup_j |\lambda_j^{-1}| < 1. \tag{2.5}$$

Any matrix A satisfying (2.5) is said to be in the Poincaré set and any matrix that does not satisfy (2.5) is said to be in the Siegel set. Note that for $n \geq 2$ dimensions the Siegel set has a non-trivial interior, and for the 1-dimensional case the set consists of $\lambda : |\lambda| = 1$. Later on, condition (2.5) was weakened to the following hyperbolic condition:

$$|\lambda_j| \neq 1 \quad \text{for} \quad 1 \leq j \leq n. \tag{2.6}$$

The last condition discussed here and the most related to small denominators is the non-resonance condition on the eigenvalues of A.

Definition 2.1.4 (Non-resonance condition)**.** Suppose $k = (k_1, k_2, \cdots, k_n) \in \mathbb{N}^n$ with $|k| = \sum k_i$, and let λ^k be $\lambda_1^{k_1}\lambda_2^{k_2}\cdots\lambda_n^{k_n}$. If

$$\lambda^k - \lambda_j \neq 0 \text{ for all } 1 \leq j \leq n \text{ and for all } k \text{ with } |k| \geq 2, \tag{2.7}$$

then A satisfies the non-resonance condition.

Earlier results

The first major result on the holomorphic linearization problem is attributed to Poincaré and Koenigs [52], and known as the Poincare-Koenigs theorem[4]. It states that if A satisfies the non-resonance condition (2.7) and condition (2.5) then f is analytically linearizable.

If the non-resonance condition (2.7) holds while (2.5) is violated, the situation becomes subtle, as small denominators begin to occur in the formal solution of h. These denominators were a known problem and remained unsolved despite the efforts of many mathematicians of the time. The problem was more or less stagnant for some years. In certain cases, it was proven that these denominators cause divergence and prevent linearization. The first example of this was given by Pfeiffer in 1917 [70]. Later, in 1938, Cremer gave a non-linearizable counterexample for every A satisfying a particular eigenvalue condition [27].

The first theorem that conclusively proved that it was possible to linearize f with A in the Siegel set, is Siegel's centre theorem, which was published in his landmark 1942 study [92]. This is commonly seen as the first adequate solution to a small denominator problem.

Small denominators

Before we continue, let us consider a particular property of the Siegel set. Let λ_j be the eigenvalues of X, and let k and λ^k be as defined in Definition 2.1.4. For $m \in \mathbb{N}$ with $m \geq 2$, we define

$$\Omega(m) = \min_{2 \leq |k| \leq m} \min_{1 \leq j \leq n} |\lambda^k - \lambda_j|. \tag{2.8}$$

If A is in the Siegel set, i.e. $\sup_j |\lambda_j| = 1$ or $\sup_j |\lambda_j^{-1}| = 1$, then we have

$$\lim_{m \to \infty} \Omega(m) = 0. \tag{2.9}$$

This issue can be particularly challenging as the terms $\lambda^k - \lambda_j$ appear as denominators in the formal solution of the conjugation h. To demonstrate this, let us consider the 1-dimensional linearization problem with $|\lambda| = 1$, and expand f and h to their respective Taylor series,

$$f = \lambda z + \sum_{n=2}^{\infty} a_n z^n \quad \text{and} \quad h = z + \sum_{n=2}^{\infty} b_n z^n. \tag{2.10}$$

[4]Some authors refer to this theorem as just the Poincaré theorem or Koenigs theorem.

The problem of finding coefficients b_n such that $h^{-1} \circ f \circ h(z) = \lambda z$ can be solved inductively, resulting in the following recurrence relation:

$$b_{n+1} = \frac{P(a_2, \cdots, a_{n+1}, b_2, \cdots, b_n)}{\lambda^{n+1} - \lambda}, \tag{2.11}$$

where P is an explicit and known multivariate complex polynomial. The linearization problem is thus reduced to the question of whether the Taylor series of h converges.

It can be proven that P_n sufficiently decays as $n \to \infty$. However, the denominators $\lambda^{n+1} - \lambda$ are more difficult. For unfortunately chosen eigenvalues[5], some of these denominators eventually become very small at some $n \in \mathbb{N}$, leading to blown-up terms in the series and complicating the convergence question.

2.1.3 Siegel's centre theorem

As mentioned before the first major breakthrough on the small denominators case of the linearization problem came from Siegel. In particular, Siegel proved the 1-dimensional case.

Theorem 2.1.5 (Siegel's centre theorem, 1-dimensional case.)**.** *Let f be*

$$f = \lambda z + O(z^2) \quad \textit{with} \quad \lambda = e^{i2\pi\omega}. \tag{2.12}$$

If ω satisfies the Diophantine conditions, then f is linearizable.

In this theorem the Diophantine conditions refer to the following variant:

Definition 2.1.6 (Diophantine conditions for 1-dimensional maps.)**.** Let $\tau, \gamma > 0$ and $\omega \in \mathbb{R}$, ω is a Diophantine number if it satisfies

$$\left| \omega - \frac{p}{q} \right| > \gamma q^{-\tau - 1} \text{ for all } \frac{p}{q} \in \mathbb{Q}. \tag{2.13}$$

The conditions for the multi-dimensional case are much more involved and are omitted from this thesis. Instead, interested readers can refer to the studies by Bruno [21] and Herman [45].

The goal of the Diophantine conditions is quite simple: they provide a lower bound for $\lambda^{n+1} - \lambda$. If ω is a Diophantine number and $\lambda = e^{i2\pi\omega}$, then for all $n \geq 2$ we have:

$$|\lambda^{n+1} - \lambda| \geq q|q|. \tag{2.14}$$

This bound is proven by considering the best rational approximation of ω as $\frac{p}{q}$. For a full proof, see the studies by Bruno [21, 22] or the well-written tutorial by Marmi [56].

Furthermore, Siegel was the first who showed that the set[6] of all λ that satisfy the condition (2.13) is non-empty. Even more remarkable, Siegel showed that this set approaches full measure as γ approaches zero, meaning that for suitable small γ almost all $\lambda \in \mathrm{GL}(\mathbb{C}, n)$ are contained within this set.

Since then, this theorem has been refined and has become a popular theorem to demonstrate improvements and new techniques concerning small denominators for some time. Some of these improvements have since been applied in KAM theorems.

2.1.4 Further developments

New solutions to the linearization problem

The linearization problem regained new interest after Siegel's theorem. While major steps were made concerning small denominators by KAM theory, several new theorems were also published for A within the Poincaré set.

In the 1950s, the (2.5) condition was weakened to (2.6). Suppose A satisfies both the non-resonance condition (2.7) and the hyperbolic condition (2.6), then f is linearizable up to C^∞ by means of Sternberg's theorem [95, 96] and its generalization by Chaperon [24].

In the 1960s, topological linearization was introduced by Grobman and Hartman, which further loosened the conditions for linearization. If A only satisfies the hyperbolic condition (2.6), and does not necessarily satisfy (2.7), then the holomorphism is topologically linearizable[7] by the Hartman-Grobman theorem [34, 41, 42]. It has been proven that topological linearization is equivalent to holomorphic linearization [24, 74], allowing the Hartman-Grobman theorem to be applied to the complex linearization problem. This is a powerful theorem and has since become the preferred way to determine if functions are linearizable.

[5] A measure-theoretic large subset of eigenvalues do in fact not result in small denominators, as was proven by Siegel.
[6] This set has since become known as the Siegel set or disk.
[7] More recently, it was proven that this linearization can be C^1-smooth, as stated in [54].

Optimal strong non-resonance conditions

In 1972, Bruno published a dual-part study [21, 22] where he proved the linearization theorem with more general condition than the Diophantine conditions for maps. These conditions, nowadays known as the Bruno condition[8] or Bruno numbers, was the first improvement of the arithmetic condition for maps. In these studies, Bruno used more or less the same methods as Siegel.

Later on, Bruno and Rüssmann generalized this condition[9] to multiple dimensions and the flow of vector fields, and it was proven to be optimal for a wide range of small denominator problems. For more information, see Bruno's book [23] and Rüssmann's studies [77–79].

In 1992, Yoccoz proved that the Bruno condition was optimal for Siegel's theorem. By studying the specific holomorphic functions, $P_\lambda = \lambda z - z^2$ where $\lambda = e^{2\pi i\alpha}$, which are the worst cases for linearizability. Yoccoz proved that the Bruno condition was in fact the most optimal for this linearization and by extension all others. Yoccoz achieved this using a novel approach, which he named geometric re-normalisation. For the original proof see Yoccoz [100], and for a more elementary explanation to small denominators and Yoccoz's proof see the introduction to small denominators written by S. Marmi [56].

Generalizations of Siegel's theorem

The generalisation of Siegel's theorem to the germs $(\mathbb{C}^n, 0)$ was made by Sternberg in [97] 1961 and later refined by Gray [33]. Later on, Siegel himself generalized his result to complex n-dimensional vector fields near a singular point [93][10].

2.1.5 Link with KAM theory

KAM theory for maps and for vector fields.

The original context of the linearization problem is a map. As mentioned before, this sets it apart from the small denominator problems for the flow of vector fields.

A significant difference between these small denominator problems is the absence of geometric considerations, such as the preservation of symplectic forms. Nevertheless, large similarities do exist between them.

Classical KAM theory is a persistence theory, and most KAM theories for maps can be interpreted in the same way. The conditions used to deal with small denominators, despite using different notation and methods, work on the same principle. The methods themselves are also similar. For example, many small denominator problems for maps can be solved and are often solved by applying the KAM method. Therefore, it is rather common to refer to these solutions using the KAM method as a KAM theory for the specific map. One notable example of this is Moser's study on twist mappings, which is one of the studies from which KAM derives its name.

It is correct to use the term KAM theory to refer to KAM theory for both maps and the flow of vector fields. However, due to the additional geometric considerations in KAM theory for vector fields, different conditions and notation, it is important to be aware of the distinction.

Besides the previous section, this thesis does not explicitly provide examples of small denominator problems and KAM theory for maps. It is instead focused on small denominators in the classical KAM setting concerning quasi-periodic tori in a geometric context.

The circle endomorphism

Finally, let us give an honorable mention to the circle endomorphism, which is perhaps the most well-understood small denominator problem and a very common example used to demonstrate the KAM method and the properties of Diophantine conditions. The circle endomorphism is defined as follows:

$$R : \mathbb{R} \times S^1 \mapsto \mathbb{R} \times S^1 : (\alpha, z) \mapsto (\alpha, e^{i2\pi\alpha} z), \tag{2.15}$$

where we are interested in the persistence of the circles $(\alpha, \cdot)$ under the circle endomorphism R. This map removes most geometric considerations from the KAM method and thus results in an easier proof.

But as a direct corollary to Siegel's linearization theorem, it has a distinctly different origin than KAM theory itself. As such, there are many alternative methods from the linearization problem that could be used to solve the problem besides the KAM method.

[8] Cf. [8].

[9] This generalized condition is referred to as the Bruno-Rüssmann condition, see section 2.2.2.

[10] Exact date is unknown to the author, the result was found in the collections of his work.

2.2 Arithmetic conditions for vectors

The exclusion of resonant frequencies alone is not sufficient for the convergence of the series. Further restrictions must be made to avoid near-resonant frequencies and to manage the magnitude of $\langle k, \omega \rangle$. This can be achieved by imposing various conditions on the frequency space, which are commonly known as arithmetic conditions.

2.2.1 Classical Diophantine conditions

The first arithmetic conditions used in KAM theory are the Diophantine conditions. These conditions were first introduced by Kolmogorov in his original 1957 study, and they remain the most commonly used arithmetic conditions in the KAM literature.

These conditions straightforwardly impose a lower limit on the absolute value of $\langle \omega, k \rangle$ for all $k \in \mathbb{Z}^n \setminus \{0\}$.

Definition 2.2.1 (Diophantine conditions)**.** Let $\omega \in \mathbb{R}^n$ be a frequency vector. If ω satisfies

$$|\langle \omega, k \rangle| \geq \frac{\gamma}{|k|^\tau} \tag{2.16}$$

for all $k \in \mathbb{Z}^n \setminus \{0\}$, where $\gamma > 0$ is the so-called Diophantine constant and $\tau > n - 1$, then ω is said to satisfy the *Diophantine* conditions for γ and τ. The subset of $\mathbb{R}^n$ of frequency vectors that satisfy the Diophantine conditions with constants γ and τ is denoted by $\mathbb{R}^n_{\gamma,\tau}$.

The set of frequencies that satisfy these conditions has some interesting properties. The set $\mathbb{R}^n_{\gamma,\tau}$ has a "Cantorized" structure (see definition 2.2.3) and positive measure, which approaches a full measure as γ decreases.

The structure of the Diophantine frequencies is important since it is often pulled back into the parameter or phase space of dynamical systems. In these cases, understanding the structure is directly linked to understanding the structure of the parameter or phase space.

The large measure property is particularly important to KAM theory since this property is the basis of the argument that almost all quasi-periodic tori survive any small perturbation.

Submanifolds Σ of the frequency space $\mathbb{R}^n$ that are sufficiently "bent" contain a large amount of Diophantine frequencies, i.e. the subset of Diophantine frequencies contained within Σ has a positive measure in Σ.

Implications to KAM theory

The large measure property of Diophantine frequencies is particularly remarkable when considered in conjunction with the Cantorized structure of the set. Let $\tau > n - 1$ and $\gamma > 0$.

The Cantorized structure implies that the set of Diophantine frequencies is topologically small, meaning that it has no interior points. At the same time the measure of $\mathbb{R}_{\gamma,\tau}$ is greater than zero. This is referred to as a *positive* or *non-trivial* measure, and is unusual for general Cantor sets. If we let γ approach zero, then the measure of the compliment $\mathbb{R}_{\gamma,\tau}$ in $\mathbb{R}^n$ approaches zero[11]. Therefore, for arbitrary γ, it is sometimes said that the Diophantine frequencies have *full* measure.

A normalized measure, such that the measure of the entire space is 1, can be interpreted as a probability measure. This implies that a randomly chosen frequency from $\mathbb{R}^n$ has a large or non-trivial[12] chance of being Diophantine, depending on the values of γ and τ.

Within KAM theory, this is often interpreted as follows: a quasi-periodic torus has a non-trivial chance, depending on the values of γ and τ, of being Diophantine. This is equivalent to stating that many quasi-periodic tori are Diophantine. Since Diophantine tori survive the perturbation, it is therefore often said that many quasi-periodic tori survive a given perturbation.

As previously mentioned, the chance of a torus being Diophantine is determined by the constants γ and τ. The constant τ is typically set to a value equal or greater than $n - 1$, which is necessary for $\mathbb{R}\gamma, \tau$ to be of positive measure. The constant γ is usually set to a small value, as this implies that the measure of $\mathbb{R}\gamma, \tau$ is relatively large. Since γ is often a factor in the size of the allowed perturbation in KAM theory, γ cannot approach zero, as it would imply that the perturbation is zero. Consequently, it cannot be claimed that the measure of surviving quasi-periodic tori is full or that almost all quasi-periodic tori survive.

Sometimes, due to a lack of parameters, KAM theory considers quasi-periodic tori that do not range over the full frequency space $\mathbb{R}^n$, but instead over a submanifold $\Sigma \subset \mathbb{R}^n$ of the frequency space. For meaningful results, this submanifold must still contain a "large" subset of Diophantine frequencies, which is achieved by requiring Σ to be sufficiently "bent".

[11] This is also true when restricted to compact subset of $\mathbb{R}^n$.
[12] This means non-zero in this context.

Structure of the set of Diophantine frequencies

The first two properties we are considering are related to the structure of Diophantine frequencies.

Proposition 2.2.2 (Properties)**.** *Let $\gamma > 0$, $\tau > n-1$, and let $\mathbb{R}^n_{\gamma,\tau}$ be the associated Diophantine frequencies. Then $\mathbb{R}^n_{\gamma,\tau}$ has the following properties.*

- *Suppose $\omega \in \mathbb{R}^n_{\gamma,\tau}$ then $c\omega \in \mathbb{R}^n_{\gamma,\tau}$ for all $c \geq 1$. i.e. $\mathbb{R}^n_{\gamma,\tau}$ is a union of closed half-lines.*
- *Let S^{n-1} be the unit ball, then $S^{n-1} \cap \mathbb{R}^n_{\gamma,\tau}$ is the union of a countable set and a Cantor set.*

Proof. The first property is obvious as for all $c > 1$ we have

$$|\langle c\omega, k\rangle| = c\,|\langle \omega, k\rangle| \geq |\langle \omega, k\rangle| \geq \frac{\gamma}{|k|^\tau}. \tag{2.17}$$

The second property is a consequence of the Cantor-Bendixson theorem (see [43]), which states that any closed subset A of a metric space with a countable base can be expressed as the disjoint union of a countable set C and a perfect set P. The perfect set P is made up of all the condensation points of A, while the countable set C consists of all the points in A that are not condensation points, meaning that each point in C has a non-trivial neighborhood that contains only a countable number of points in A.

It is well-known that S^{n-1}, as subset of $\mathbb{R}^n$, is a second-countable space, and since $\mathbb{R}^n_{\gamma,\tau}$ is closed, so is $S^{n-1} \cap \mathbb{R}^n_{\gamma,\tau}$. Therefore we can apply the Cantor-Bendixson theorem. As $S^{n-1} \cap \mathbb{R}^n_{\gamma,\tau}$ only consists of condensation points, the perfect set P is in fact all of $S^{n-1} \cap \mathbb{R}^n_{\gamma,\tau}$. The countable set C is the empty set, since $S^{n-1} \cap \mathbb{R}^n_{\gamma,\tau}$ is dense in $S^{n-1} \cap \mathbb{R}^n$.

Let us now consider the hyperplanes defined by $\omega \in S^{n-1} : \langle \omega, k\rangle = 0$ that densely fill the complement of $S^{n-1} \cap \mathbb{R}^n_{\gamma,\tau}$ within S^{n-1}. Each of these hyperplanes separates S^{n-1} into two disconnected sets, and since these hyperplanes are dense in S^{n-1}, it is always possible to find a hyperplane that separates any two points in $S^{n-1} \cap \mathbb{R}^n_{\gamma,\tau}$. Therefore, $S^{n-1} \cap \mathbb{R}^n_{\gamma,\tau}$ is totally disconnected.

Since a totally disconnected perfect set is, by definition, a Cantor set, and is homeomorphic to the traditional Cantor set, we can conclude that $S^{n-1} \cap \mathbb{R}^n_{\gamma,\tau}$ is a Cantor set.

□

The first property is quite useful as it allows us to reduce many problems involving Diophantine frequencies to the $\gamma = 1$ case. This is achieved by first considering $\mathbb{R}^n_{\gamma,\tau}$ and re-scaling the frequencies from ω to $\gamma^{-1}\omega$ through an appropriate time re-scaling, thus obtaining $\mathbb{R}^n_{1,\tau}$. As γ is small and $\gamma^{-1} > 1$, we can use the first property to conclude that proving any theorem with $\gamma = 1$ is equivalent to proving it for smaller γ.

The second property is not often used directly, but in conjunction with the first property, it provides valuable insight into the overall structure of the set of Diophantine frequencies.

Together, the first and second properties imply that the set of Diophantine frequencies can be interpreted as a foliation of the half lines $c\omega : c \in \mathbb{R}$ parameterized by ω in the Cantor set $S^{n-1} \cap \mathbb{R}^n_{\gamma,\tau}$. Furthermore, this foliation is smooth in the sense of Whitney (see section 2.3). This allows us to formalize what is understood by a Cantorized structure:

Definition 2.2.3 (Cantorized structure)**.** A Whitney-smooth foliation of closed half lines parameterized by a Cantor set is colloquially referred to as a *Cantorized* set.

A set with a Cantorized structure has gaps resulting from the gaps naturally present in the Cantor set. In the Diophantine frequency set, these gaps arise from the exclusion of resonances and their associated near-resonances.

We define the following.

Definition 2.2.4 (Resonance gaps)**.** The gaps left by excluding each resonance and its associated near-resonances are called *resonance gaps*.

Measure of the set of Diophantine frequencies

The third and perhaps the most important property of the set of Diophantine frequencies is that its complement in $\mathbb{R}^n$ approaches zero measure as γ approaches zero. This property makes the set of Diophantine frequencies quite exceptional, as Cantor sets typically have a trivial measure.

Lemma 2.2.5 (Positive measure)**.** *Let S^n be the unit sphere and let $|\cdot|_\lambda$ be the Lebesgue measure normalized such that $\mu(S^n) = 1$. Lastly let $\tau > n-1$. Then the Lebesgue measure of $S^n \setminus S^n_{\gamma,\tau}$ is of the order γ as γ approaches zero, where $S^n_{\gamma,\tau} = S^n \cap \mathbb{R}^n_{\gamma,\tau}$.*

Proof. Let us define the set of frequencies that fail the Diophantine condition for a fixed $k \in \mathbb{Z}^n \setminus \{0\}$ as

$$\mathcal{B}_k = \left\{\omega \in \mathbb{R}^n \,:\, |\langle \omega, k\rangle| < \gamma |k|^{-\tau}\right\}. \tag{2.18}$$

With $\mathcal{B}_k$, the set of all Diophantine frequencies can be written as

$$\mathbb{R}^n_{\gamma,\tau} = \mathbb{R}^n \setminus \left(\bigcup_{k\in\mathbb{Z}^n\setminus\{0\}} \mathcal{B}_k \right). \tag{2.19}$$

Let us now consider the measure of $\mathcal{B}_k$. Geometrically, $\mathcal{B}_k$ is a strip[13], and the thickness of each $\mathcal{B}_k$ is at most $\gamma|k|^{-\tau-1}$. Therefore, the measure of $\mathcal{B}_k \cap S^n$ is bounded by $\gamma|k|^{-\tau-1}$. Consequently, we can calculate the measure of $S^n \setminus S^n_{\gamma,\tau}$ as follows

$$\left|S^n \setminus S^n_{\gamma,\tau}\right|_\lambda = \left| \bigcup_{k\in\mathbb{Z}^n\setminus\{0\}} S^n \cap \mathcal{B}_k \right|_\lambda \leq \sum_{k\in\mathbb{Z}^n\setminus\{0\}} |S^n \cap \mathcal{B}_k|_\lambda \leq \gamma \sum_{k\in\mathbb{Z}^n\setminus\{0\}} |k|^{-\tau-1}. \tag{2.20}$$

To determine when the series $\sum_{k\in\mathbb{Z}^n\setminus\{0\}} |k|^{-\tau-1}$ converges, we substitute $|k|$ using the following estimate of the number of k that satisfy $|k| = l$

$$\#\{k \in \mathbb{Z}^n : |k| = l\} \leq 2^n l^{n-1}. \tag{2.21}$$

It follows that

$$\sum_{k\in\mathbb{Z}^n\setminus\{0\}} |k|^{-\tau-1} \leq 2^n \sum_{l=1}^{\infty} l^{-\tau-1} l^{n-1} = 2^n \sum_{l=1}^{\infty} l^{n-\tau-2}, \tag{2.22}$$

which converges if $n - \tau - 2 < -1$ or equivalently if $\tau > n-1$. Thus, we conclude that

$$\left|S^n \setminus S^n_{\gamma,\tau}\right|_\lambda \leq \gamma 2^n \sum_{l=1}^{\infty} l^{n-\tau}. \tag{2.23}$$

□

The importance of this result cannot be overstated.

From the above result, where the measure of the complement approaches zero, it follows that the measure of the set of Diophantine frequencies approaches the full measure. This implies that almost all frequencies satisfy the Diophantine conditions for small enough γ.

As Diophantine conditions are the main conditions for the persistence of quasi-periodic tori, it follows that many quasi-periodic tori survive small perturbations. This has significant dynamical implications.

The success of KAM theory depends on the fact that many tori survive a small perturbation, as otherwise, it would arguably not be a very useful persistence theory. A KAM theorem that proves that a measure-theoretic negligible amount of quasi-periodic tori survive has little dynamical meaning, and the incidental persistence of quasi-periodic tori can be readily demonstrated with specifically chosen examples and perturbations.

[13] A hyper-surface determined by the direction of k times a small interval.

Submanifolds of the set of Diophantine frequencies

In this section, we present two theorems that generalize the full measure argument discussed above to sub-manifolds of the frequency space $\mathbb{R}^n$, without providing their proofs.

These theorems argue that, within lower dimensional sub-manifolds of the frequency space, the subset of Diophantine frequencies in these sub-manifolds have full measure as γ approaches zero. Studying lower dimensional manifolds of the frequency space instead of the full frequency space eases the amount of variables required to describe all frequencies. This is beneficial, as this directly decreases the amount of parameters required to control the frequencies (see section 3.5). These techniques are known as Diophantine approximation of dependent quantities and is applicable to both sub-manifolds and images of mappings.

Pyartli's theorem

We start with the Diophantine approximation on sub-manifolds $\Sigma \subset \mathbb{R}^n$. When this manifold is sufficiently "bent" or "curved", the subset of Diophantine frequencies contained within it has a positive measure. This argument is sometime referred to as the "curvature" argument, however this is somewhat of a misdirection as it is not related to what is commonly understood by curvature, i.e. curvature as defined on Riemannian manifolds.

Let Σ be a d-dimensional compact C^{l+1} submanifold of $\mathbb{R}^n$, which is defined by the C^{l+1} immersion map $\omega : W \to \mathbb{R}^n$, where $W \subset \mathbb{R}^d$ is a compact domain. The manifold Σ is sufficiently bent if, for every point $x \in \Sigma$ of the submanifold, we have that

$$T_x\Sigma + T_x^2\Sigma + \cdots + T_x^l\Sigma = T_x\mathbb{R}^n, \tag{2.24}$$

where T^j is the j-th tangent plane of Σ, defined by $T^j\Sigma = D^j\omega$.

With this, we can state the following theorem

Theorem 2.2.6 (Pyartli)**.** *Let Σ be as above and sufficiently bent, then for $\tau > n^2 + n + 1$ there exists a $\gamma = \gamma(\tau) > 0$ such that almost every point $w \in W$ satisfies*

$$|\langle \omega(w), k\rangle| \geq \gamma |k|^{-\tau}. \tag{2.25}$$

The proof for this theorem can be found in the original study by Pyartli [75].

Bakhtin's theorem

The argument of Pyartli was generalized by Bakhtin in 1991 to include general frequency mappings, which are more flexible. This generalization extends Pyartli's theorem, as Bakhtin's theorem can be applied to all immersion mappings defining submanifolds.

We start by defining the frequency mapping. Let $W \subset \mathbb{R}^d$ be an open domain, $\omega : W \to \mathbb{R}^n$ be an analytical function, and $\Gamma \subset W$ be diffeomorphic to the closed unit sphere in $\mathbb{R}^d$. Lastly, let us denote

$$D^q\omega := \frac{\partial^{|q|}\omega(w)}{\partial w^q} \quad \text{with} \quad q \in \mathbb{Z}_+ \quad \text{and} \quad w^q := w^{q_1} w^{q_2} \cdots w^{q_d}. \tag{2.26}$$

With this, we can state the following theorem

Theorem 2.2.7 (Bakhtin)**.** *Let W, Γ and $\omega : W \to \mathbb{R}^n$ be as above. Suppose $w \in \Gamma$ and that the collection of $(d+Q)!/d!Q!$ vectors defined by*

$$D^q\omega(w),\ q \in \mathbb{Z}_+,\ 0 \leq |q| \leq Q \tag{2.27}$$

spans $\mathbb{R}^n$. Then for $\tau \geq nQ - 1$ and $\varepsilon^* > 0$ there exists a $\gamma = \gamma(\tau, \varepsilon^*) > 0$ such that the Lebesgue measure of the subset of Γ for which

$$|\langle \omega(w), k\rangle| \geq \gamma |k|^{-\tau} \tag{2.28}$$

holds is greater than $(1 - \varepsilon^*)\, meas_d\, \Gamma$.

Again, we omit the proof here for the sake of brevity. For a detailed proof of the theorem, please refer to the original paper by Bakhtin [5]. Alternatively, for a similar proof that is applicable to normal-internal frequencies, please see the appendix of the book by Broer, Huitema, and Sevryuk [17].

2.2.2 Generalized arithmetic conditions

Since their initial inception by Siegel, the Diophantine conditions have been generalized. These generalizations can be expressed using the approximation function formulation, as described below.

Approximation function formulation

The approximation function formulation of arithmetic conditions is heavily inspired by the Bruno condition (see [21, 22]). This condition was adapted by Rüssmann [77, 78] to create the Bruno-Rüssmann condition, which also uses the approximation function formulation.

Definition 2.2.8 (Approximation function)**.** Let $\Delta : [1, \infty) \to \Delta([1, \infty)) \subset \mathbb{R}$ be an arbitrary monotonically increasing continuous function such that $\Delta(1) > 0$ and

$$\int_1^\infty \frac{\log \Delta(u)}{u^2}\, du < \infty, \tag{2.29}$$

then $\Delta(u)$ is an *approximation function*.

By using an approximation function, we can express more general arithmetic conditions as

$$|\langle \omega, k \rangle| \geq \frac{\gamma}{\Delta(|k|)} \quad \text{for all} \quad k \in \mathbb{Z}^n \setminus \{0\}. \tag{2.30}$$

These conditions are similar to the Diophantine conditions, as the frequencies that satisfy the approximation conditions have a positive measure. In fact, for most well-chosen approximation functions, we can easily prove that the Diophantine frequencies are contained within the frequencies that satisfy condition (2.30). For a complete description of these conditions, see Pöschel [72] and Rüssmann [80].

If we set $\Delta(u) = u^\tau$, we can obtain the original Diophantine conditions from the approximation function formulation.

Bruno-Rüssmann condition

The Bruno-Rüssmann is the most optimal[14] arithmetic condition so far. This condition has been co-developed by Bruno [21, 22] and Rüssmann [77, 78] for the Schröder-Siegel problem.

Definition 2.2.9 (Bruno-Rüssmann condition)**.** $\omega \in \mathbb{R}^n$ satisfies the Bruno-Rüssmann condition if it satisfies (2.30) with

$$\Delta(u) = \max\{|\langle \omega, k \rangle|^{-1} : k \in \mathbb{Z}^n \text{ with } 1 \leq |k| \leq u\}. \tag{2.31}$$

In one dimension, this reduces to the Bruno condition, which, as mentioned before, has been proven optimal for the linearization of the standard map in one dimension by Yoccoz [100, 101]. (See the previous section for more information.)

This condition has also been proven optimal for several problems in $\mathbb{C}^2$, notably the complex area-preserving map known as the semi-standard map [55] and for vector fields close to a non-resonant singular point [69]. Here, it is usually stated in terms of Bruno numbers, where $\omega = (1, \alpha)$ and α is a Bruno number.

The optimality of the Bruno-Rüssmann condition has not been determined in higher dimensions and remains an open question, as noted in [8].

Use of generalized arithmetic conditions

The vast majority of KAM theorems and small denominator problems use the classical Diophantine conditions. Further refinements are rarely necessary, as most of the time, Diophantine frequencies already have a large measure. The Bruno condition and other optimized arithmetic conditions are rarely used outside of studies concerned with optimization.

A notable exception is the use of adapted arithmetic conditions for lower-dimensional tori, as the classical Diophantine conditions as described in (2.16) are not sufficient. The classical Diophantine conditions only exclude internal near-resonances. In addition to internal near-resonances, normal-internal near-resonances also need to be excluded, as shown in the study by Pöschel [72].

Therefore, slightly modified arithmetic conditions must be used. These conditions are very similar in form and function to the classical Diophantine conditions and are also commonly referred to as Diophantine conditions.

[14] To the best knowledge of the author.

2.3 Whitney regularity

It is well-known that functions form spaces under composition which preserve properties such as continuity and analyticity. It is common to say that a function is contained within the space of all analytic or continuous functions.

These spaces can be extended to a sliding scale. For example, the space of continuous functions C^0 has subspaces C^k based upon continuity of the derivative. Each higher subspace C^{k+1} is contained in the lesser subspace C^k, making it a sliding scale. Functions are usually classified based on membership of the smallest space in the scale that still contains them. The membership of a function to a certain space is referred to as the regularity of a function. Within modern analysis, this concept has been further generalized by introducing sliding scales of spaces for more abstract and exotic properties of functions preserved under composition.

The most well-known regularity is the scale of spaces of a-differentiable functions, denoted by C^a, which includes the space of fractional differentiable functions. These spaces are parameterized by $a \in \mathbb{R}^+$, where each C^a is contained within C^b for all $a > b$ in $\mathbb{R}^+$, thus forming the scale of spaces. We can determine how regular a function is based on the smallest function space C^a it is contained within, and we can say that the function is C^a regular.

Another example and the main subject of this section is Whitney regularity, where functions are said to be Whitney-C^β or C^β differentiable in the Whitney sense. This regularity can be defined for functions on closed sets, and this was the main motivation for it.

The last example given is Gevrey regularity, where functions are said to be α-Gevrey. This is a scale of function spaces that exist between analytic and smooth functions. This regularity class is sometimes used instead of Whitney regularity in KAM theory.

Determining the regularity of a function is useful for many reasons. For example, each regularity could have their own specific estimates for the norms of the Fourier coefficients and other norms. Approximation lemmas and smoothing operators can also be specifically engineered for a regularity class. More importantly, for each more regular function space, most of the estimates and approximations can be made sharper. Some theorems are dependent on the functions having a minimum regularity.

2.3.1 Situation within small denominator problems

The solution to many small denominator problems often results in a formal series, denoted by h_ω, that depends on a frequency ω. We can interpret the frequency ω as a parameter within a parameter space $\Gamma \subset \mathbb{R}^n$. In such situations, it can be advantageous to study the regularity of h_ω in the ω direction.

The main complication is that h_ω is only well-defined for specific Diophantine frequencies in $\Gamma_\gamma = \Gamma \cap \mathbb{R}^n_{\gamma,\tau}$. This set does not have any interior points, and as the most common regularities rely on the existence of an interior in their definition, studying regularity in this context is complicated. This requires a different interpretation of regularity that is applicable to these types of domains.

One possible type of regularity that can be defined for such sets is Lipschitz continuity. This is a regularity class that can be readily defined on closed sets and thus to the Γ_γ set. This is demonstrated by Pöschel in his lecture notes [73]. However, Lipschitz continuity is a relatively weak regularity, and stronger alternatives are available.

2.3.2 Regularity of functions in arbitrary closed sets

It is possible to study the regularity of a function on a closed set by extending[15] the function to a neighborhood of the closed set. On this neighborhood, the usual definition of regularity classes can be applied.

This is achieved through the Whitney extension theorem and the associated definition of Whitney differentiability. Roughly speaking, a function that is smooth in the Whitney sense on a closed set can be extended to a smooth function defined on an open neighborhood of that closed set or even the entire underlying space.

The Whitney extension theorem[16] has been adapted for many regularity classes, for example Gevrey-regularity. However, since all examples and theorems given in this thesis are for Whitney-smooth differentiability[17], only this regularity is explained.

The study of regularity in the ω-direction in the Whitney sense is now a common part of KAM theory, and therefore Whitney regularity is considered in the demonstration of the one-bite small denominator problem and the KAM proof given in this thesis.

[15] With respect suitable regularity.
[16] And associated Whitney-regularity definitions and approximation lemma.
[17] Whitney-C^β differentiability with β arbitrarily high.

2.3.3 Definition of Whitney-C^β differentiability.

Whitney-C^β differentiability is both defined on open and closed domains. We consider the Whitney-C^β differentiability on the subset $\Omega \subset \mathbb{R}^n$, where separate definitions are given for open and closed Ω. In case that $\Omega = \mathbb{R}^n$, which is both open and closed, the definitions are equivalent.

We begin with the open domain definition and then continue with the closed set definition. These definitions are quoted from Whitney's original study [99].

But first, we discuss some notation. Let $\beta > 0$ be a fixed order of differentiability and define l to be the largest positive integer such that l is strictly lower than β, i.e.

$$l := \max\{m \in \mathbb{Z} : m < \beta\}. \tag{2.32}$$

The following sections also use the shorthand $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$.

Open sets

Definition 2.3.1 (Whitney-C^β differentiability)**.** Let $\Omega \subset \mathbb{R}^n$ be an open domain.

The class of $C^\beta_{\text{Wh}}(\Omega)$ of Whitney-C^β functions in Ω consists of C^l functions $U : \Omega \to \mathbb{R}$ with uniformly bounded derivatives up to order l, and where the l-th derivative satisfies a Hölder condition with exponent $\beta - l$.

In other words, for all $U \in C^\beta_{\text{Wh}}(\Omega)$ there exists an $M > 0$ such that for all $x \in \Omega$

$$|D^q U(x)| < M \text{ for all } q \in \mathbb{Z}^n_+ \text{ with } 0 \leq |q| \leq l, \tag{2.33}$$

and for all $x, y \in \Omega$

$$|D^q U(x) - D^q U(y)| \leq M|x - y|^{\beta - l} \text{ for } q \in \mathbb{Z}^n_+ \text{ with } |q| = l. \tag{2.34}$$

Lastly the norm $||U||_\beta$ on $C^\beta_{\text{Wh}}(\Omega)$ is the C^β-Hölder norm, i.e. the smallest M for which (2.33) and (2.34) hold.

Closed sets

Definition 2.3.2 (Whitney-C^β differentiability)**.** Let $\Omega \subset \mathbb{R}^n$ be a closed subset[18].

The class of $C^\beta_{\text{Wh}}(\Omega)$ of Whitney-C^β functions in Ω consists of collections $U = \{U_q\}_{0 \leq |q| \leq l}$ with $q \in \mathbb{Z}^n_+$ of functions defined on Ω for which the following holds.

There exists an $M > 0$ such that for all $x, y \in \Omega$

$$|U_q(x)| \leq M \text{ for all } q \in \mathbb{Z}^n_+ \text{ with } 0 \leq |q| \leq l, \tag{2.35}$$

and

$$|U_q(x) - P_q(x, y)| \leq M|x - y|^{\beta - |q|} \text{ for } q \in \mathbb{Z}^n_+ \text{ with } |q| = l, \tag{2.36}$$

where

$$P_q(x, y) = \sum_{j=0}^{l-|q|} \sum_{|k|=j} \frac{1}{k!} U_{q+k}(y)(x - y)^k.$$

Lastly the norm $||U||_\beta$ on the collections $U \in C^\beta_{\text{Wh}}(\Omega)$ is defined as the smallest M for which (2.35) holds.

[18] For $\Omega = \mathbb{R}^n$ the open domain definition suffices.

2.3.4 The lemma and theorem

The following lemma and theorem are advanced and therefore their proofs are not supplied, but only referenced, see [71, 94, 99, 103]. In the following paragraphs, the shorthand $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$ is again used.

Whitney inverse approximation lemma

The definition of Whitney-C^β regularity can be difficult to verify for the functions one may want to determine the regularity of.

The typical method to determine the Whitney differentiability of a function is to define a sequence of geometrically shrinking domains W_j and functions $U^j : W_j \to \mathbb{R}$ that respectively approach the function $U^\infty : \Omega \to \mathbb{R}$ and the domain Ω in question. This spares us from having to define the function U^∞ on closed Ω ourselves, and in most cases, U^∞ is only available as the limit of a sequence.

This method uses the Whitney inverse approximation lemma, which can be used to determine the Whitney-C^β regularity of functions defined as the limit of a sequence of functions. It requires the following set-up;

Suppose we have an open or closed subset $\Omega \subset \mathbb{R}^n$.

Let $r_j = \alpha\kappa^j$, $j \in \mathbb{Z}_+$ be a geometrically shrinking sequence with $\alpha = r_0 > 0$ and $0 < \kappa < 1$, and consider the sequence of complex domains $W_j = \Omega + r_j$, defined by

$$W_j = \bigcup_{x\in\Omega} \{z \in \mathbb{C}^n \ : \ |z - x| < r_j\}. \tag{2.37}$$

Lastly define $\{U_j\}_{j\geq 0}$, where each $U_j : W_j \to \mathbb{R}$ is an analytic function and $U_0 \equiv 0$.

The inverse approximation lemma answers the question; what is the C^β-Whitney class of the limit of $\{U_j\}_{j\geq 0}$, denoted by U_∞.

Lemma 2.3.3 (Inverse approximation)**.** *Let $\beta \in \mathbb{R}_+ \setminus \mathbb{Z}_+$, and let Ω, $\{r_j\}_{j\geq 0}$, $\{W_j\}_{j\geq 0}$ and $\{U_j\}_{j\geq 0}$ be as above.*

If for every U_j with $j \geq 1$ there exists a constant $M > 0$ such that

$$|U_j - U_{j-1}|_{W_j} \leq M r_j^\beta, \tag{2.38}$$

where $|\cdot|_{W_j}$ is the norm defined on W_j, then the limit U_∞, defined on Ω, exists and the following is true.

- *U_∞ is C^β-Whitney regular.*
- *$||U_\infty||_\beta \leq M c_{\kappa,\beta}$, where $||\cdot||_\beta$ is the C^β-Whitney norm and the constant $c_{\kappa,\beta} > 0$ only depends on κ, β and n.*
- *For all $\alpha < \beta$, $||U_\infty - U_j||_\alpha \to 0$ as j goes to infinity, where $||\cdot||_\alpha$ is the C^α-Whitney norm.*

If Ω is closed, then U_ω should be treated as the collection $\{U_q^\infty - D_q U_j\}_q$.

This lemma is extensively used within KAM literature to determine the regularity of the resulting conjugation. As mentioned before, variants of this lemma exist for other regularity classes in the sense of Whitney.

Whitney extension theorem

As was mentioned before, a Whitney-C^β function defined on a closed domain in $\mathbb{R}^n$ can be extended to the entire space of $\mathbb{R}^n$.

Theorem 2.3.4 (Whitney Extension Theorem)**.** *Suppose $\beta > 0$ is arbitrary, and $\Omega \subset \mathbb{R}^n$ is a closed subset. Then there exists a linear extension operator $\mathcal{E} : C^\beta_{Wh}(\Omega) \to C^\beta_{Wh}(\mathbb{R}^n)$ such that, for each $U = \{U_q\}_q \in C^\beta_{Wh}(\Omega)$ and $0 \leq |q| \leq l$ ($q \in \mathbb{Z}^n_+$), the following holds*

$$D^q(\mathcal{E}U)|_\Omega = U_q \quad with \quad ||\mathcal{E}U||_\beta \leq c_\beta ||U||_\beta,$$

where $||\cdot||_\beta$ is the C^β-norm over respectively $C^\beta_{Wh}(\mathbb{R}^n)$ and $C^\beta_{Wh}(\Omega)$, and c_β is a constant dependent only on n and β.

Note that here, β can indeed be any positive real number.

2.4 One-bite small denominator problem

Small denominators can arise in various forms in questions related to conjugation and linearization.

In this section, we only consider the one-bite small denominator problem. This type of small denominator problem is commonplace, as solving multiple one-bite small denominator problems can provide solutions to more complex nonlinear small denominator problems. Moreover, the one-bite problem is comparatively easy to solve as it has a linear character, and thus serves as a good example.

The variant presented below is a simple application of the Diophantine conditions and the Whitney approximation lemma. This section draws heavily on the examples and proofs given in [17] and [20].

Problem 2.4.1. Consider a linear partial differential equation of the following form

$$\left\langle \omega, \frac{\partial u}{\partial x} \right\rangle = f, \tag{2.39}$$

with $f = f(\omega, x)$ with $x \in T^n$, $\omega \in \mathbb{R}^n$. Can we solve for $u(\omega, x)$?

We directly solve for u by means of the Fourier series expansion for u and f.

2.4.1 Formal solution

Denote the Fourier series of f and u as

$$f(x,\omega) = \sum_{k\in\mathbb{Z}^n} f_k(\omega) e^{i\langle x,k\rangle} \quad \text{and} \quad u(x,\omega) = \sum_{k\in\mathbb{Z}^n} u_k(\omega) e^{i\langle x,k\rangle}. \tag{2.40}$$

We see that

$$\frac{\partial}{\partial x} u(x,\omega) = \sum_{k\in\mathbb{Z}^n} u_k(\omega) \frac{\partial}{\partial x} e^{i\langle x,k\rangle} = \sum_{k\in\mathbb{Z}^n\setminus\{0\}} iku_k(\omega) e^{i\langle x,k\rangle}, \tag{2.41}$$

where $k \in \mathbb{Z}^n \setminus \{0\}$ is an integer vector. Therefore

$$\left\langle \omega, \frac{\partial u}{\partial x} \right\rangle = \sum_{k\in\mathbb{Z}^n\setminus\{0\}} i\langle \omega, k\rangle u_k(\omega) e^{i\langle x,k\rangle} = \sum_{k\in\mathbb{Z}^n} f_k(\omega) e^{i\langle x,k\rangle}. \tag{2.42}$$

We see that this equation only has a solution if $f_0(\omega) \equiv 0$. Therefore

$$f_0(\omega) = [f(\cdot,\omega)] = \int_{T^n} f(x,\omega)dx \equiv_\omega 0, \tag{2.43}$$

or in words the T^n-average of $f(x,\omega)$ must vanish identically with respect to ω. This leads to the formal solution

$$u(x,\omega) = u_0(\omega) + \sum_{k\in\mathbb{Z}^n\setminus\{0\}} \frac{f_k(\omega)}{i\langle \omega, k\rangle} e^{i\langle x,k\rangle}. \tag{2.44}$$

Note that $u_0(\omega)$ is arbitrary and is usually set to $u_0(\omega) \equiv 0$ to ensure a unique solution.

This is a formal solution if all resonances $\langle \omega, k\rangle = 0$ are excluded, as these resonances would result in division by zero and the sum would no longer be well-defined. What remains to be done is to prove the convergence of this series under Diophantine conditions and determine its resulting regularity.

2.4.2 Whitney-smooth solution to the 1-bite problem

We prove the following theorem.

Theorem 2.4.2 (analytic 1-bite problem)**.** *Let $f : T^n \times \mathbb{R}^n \to \mathbb{R}$ be analytic in both x and ω, $[f(\cdot,\omega)] \equiv_\omega 0$, and bounded on $T^n \times \mathbb{R}^n$. Then there exists a function $u : T^n \times \mathbb{R}^n \to \mathbb{R}$ such that the restriction of u to $T^n \times \mathbb{R}^n_{\gamma,\tau}$, with $\gamma > 0$ and $\tau > n-1$, satisfies*

$$\left\langle \omega, \frac{\partial u}{\partial x} \right\rangle = f, \tag{2.45}$$

and such that u is analytic in x and Whitney-smooth in ω, as (x,ω) varies over $T^n \times \mathbb{R}^n_{\gamma,\tau}$.

Proof of theorem 2.4.2

Initial set-up and notations

To begin, let $\Gamma \subset \mathbb{R}^n$ be a compact domain. Most KAM theories are defined on a compact subset of the frequency space. However, as long as f remains uniformly bounded, this subset is arbitrary and could be the entire frequency space $\mathbb{R}^n$. The compactness of Γ ensures this uniform bound.

To control the small denominators, we restrict the frequencies to Diophantine frequencies $\mathbb{R}^n_{\tau,\gamma}$. Generally, τ is fixed to a constant higher than $n-1$, as this implies that the Diophantine frequencies have a non-trivial measure. We define $\Gamma_\gamma = \Gamma \cap \mathbb{R}^n_{\gamma,\tau}$.

We define the following open complex neighborhoods of Γ and T^n:

$$\Gamma + \rho = \bigcup_{\omega\in\Gamma} \{\hat{\omega} \in \mathbb{C}^n \,:\, |\hat{\omega} - \omega| < \rho\} \tag{2.46}$$

$$T^n + \kappa = \bigcup_{x\in T^n} \{\hat{x} \in (\mathbb{C}/2\pi\mathbb{Z})^n \,:\, |\hat{x} - x| < \kappa\}\,. \tag{2.47}$$

Since f is analytical, it has a holomorphic extension to a complex neighborhood U of $T^n \times \Gamma$ in $(\mathbb{C}/2\pi\mathbb{Z})^n \times \mathbb{C}^n$. As U has a non-trivial interior, there exist two small constants ρ and $\kappa > 0$ such that U contains the closure of $(T^n + \kappa) \times (\Gamma + \rho)$. Lastly, we define M as the supremum of the extension of f over the closure of $(T^n + \kappa) \times (\Gamma + \rho)$, i.e.

$$M = \sup_{(x,y)\in \mathrm{cl}((T^n+\kappa)\times(\Gamma+\rho))} |f(x,\omega)|. \tag{2.48}$$

Optional reduction to $\gamma = 1$ case.

In certain cases where γ is assumed to be fixed, we can further simplify the following arguments by reducing it to the $\gamma = 1$ case, by scaling the time $t \mapsto \gamma t$ and the frequencies $\omega \mapsto \gamma^{-1}\omega$. This transformation maps $\Gamma \mapsto \gamma^{-1}\Gamma$.

This scaling would require us to assume Γ is unbounded in this proof, and by extension that f is uniformly bounded. As usually only bounded perturbations are considered in KAM theory, it is relatively common to find that f is bounded. This is therefore a common reduction found in KAM proofs.

In this proof, we do not make this assumption and must keep track of γ explicitly.

Paley-Wiener lemma

As we have seen in section 2.4.1, our main concern is the convergence of the formal solution. To prove this, we begin by providing an upper bound for the Fourier coefficients of f, which can be easily achieved by applying the classical Paley-Wiener lemma.

Lemma 2.4.3 (Paley-Wiener)**.** *Let $f = f(x,\omega)$ be as in theorem 2.4.2, with Fourier series*

$$f(x,\omega) = \sum_{k\in\mathbb{Z}^n} f_k(\omega) e^{i\langle x,k\rangle}. \tag{2.49}$$

Then, for all $k \in \mathbb{Z}^n$ and $\omega \in cl(\Gamma + \rho)$, we have

$$|f_k(\omega)| \le M e^{-\kappa|k|}. \tag{2.50}$$

where M is as defined in (2.48)*.*

Proof. By definition of the Fourier coefficient

$$f_k(\omega) = \frac{1}{(2\pi)^n} \int_{T^n} f(x,\omega) e^{-i\langle k,x\rangle}\, dx. \tag{2.51}$$

For the moment assume that $k = (k_1, \cdots, k_n)$ does not have zero entry, i.e. $k_j \neq 0$ for $1 \le j \le n$, and define $z_j = x_j - i\kappa \cdot \mathrm{sign}(k_j)$. Then use Cauchy theorem's to obtain

$$f_k(\omega) = \frac{1}{(2\pi)^n} \int_{T^n} f(z,\omega) e^{-i\langle k,z\rangle}\, dz = \frac{1}{(2\pi)^n} \int_{T^n} f(z,\omega) e^{-i\langle k,x\rangle - \kappa|k|}\, dx. \tag{2.52}$$

This easily gives us

$$|f_k(\omega)| = \left| \frac{1}{(2\pi)^n} \int_{T^n} f(z,\omega) e^{-i\langle k,x,\rangle - \kappa|k|}\, dx \right| \le \int_{T^n} |f(z,\omega)| \cdot \left| e^{-i\langle k,x,\rangle - \kappa|k|} \right| dx = \int_{T^n} M e^{-\kappa|k|}\, dx = M e^{-\kappa|k|}, \tag{2.53}$$

where we used $\left|e^{a+ib}\right| = |e^a|$. Lastly suppose there are entries k_j that are equal to zero, then we just forgo the torus shift in those directions. □

The convergence of the formal solution for a fixed $\omega \in \Gamma_\gamma$ can be proven using the Paley-Wiener lemma. The lemma implies that the f_k decay exponentially, while $\omega \in \Gamma_\gamma$ implies that all $i\langle k, \omega\rangle$ are uniformly bounded from below for all $k \in \mathbb{Z}^n \setminus \{0\}$. Therefore, for a fixed ω, the formal solution for u converges.

This argument proves that for $\omega \in \Gamma_\gamma$, the formal solution (2.44) converges, implying the existence of u. However, it does not prove the regularity of u in the ω-direction, thus further argumentation is needed.

Sequence with Whitney-C^∞ limit

Our strategy is to apply the Whitney inverse approximation lemma 2.3.3 to determine the regularity of $u(x, \omega)$.

The inverse approximation lemma requires that the functions h^i and h^{i+1} in the sequence of real analytic functions $\{h^j\}_{j\geq 1}$ defined on the sequence of geometrically shrinking domains W_j, have a geometrically decaying point-wise difference as j goes to infinity.

In the case of the one-bite problem, we define the sequence, $\{u^j\}_{j\geq 0}$, by the following truncations of u

$$u^j(x, \omega) = \sum_{0<|k|\leq 2^j} \frac{f_k(\omega)}{i\langle \omega, k\rangle} e^{i\langle x,k\rangle}, \tag{2.54}$$

which uniformly converge to u.

For the shrinking domains W_j, we set $W_j = (T^n + \kappa) \times \Gamma_\gamma + r_j$, where $r_j = \gamma \cdot 2^{-j\tau-j-1}$, which geometrically shrinks to $W_\infty = (T^n + \kappa) \times \Gamma_\gamma$ as j approaches infinity.

Note that $\Gamma_\gamma + r_j \subset \Gamma + \rho$ for sufficiently large j, making each u^j well-defined on W_j. We also need to verify that the truncations u^j are holomorphic functions on the chosen domain $(T^n + \kappa) \times W_j$. This is shown by the following lemma.

Lemma 2.4.4 (Small denominator lemma)**.** *For all $\omega \in \Gamma_\gamma + r_j$ and $k \in \mathbb{Z}^n$ with $0 < |k| \leq 2^j$ one has $|\langle \omega, k\rangle| \geq \frac{1}{2}\gamma|k|^{-\tau}$.*

Proof. Suppose $\omega \in \Gamma_\gamma + r_j$, then by definition there exists an $\hat{\omega} \in \Gamma_\gamma$ such that $|\omega - \hat{\omega}| < r_j$. Therefore for all $0 < |k| \leq 2^j$, we have

$$|\langle \omega, k\rangle| \geq |\langle \hat{\omega}, k\rangle| - |\omega - \hat{\omega}| \cdot |k| \geq \gamma|k|^{-\tau} - r_j|k|.$$

Now as $|k| \leq 2^j$, we state

$$r_j|k| \leq \gamma 2^{-j\tau-j-1}2^j = \tfrac{1}{2}\gamma \cdot 2^{-j\tau} = \tfrac{1}{2}\gamma(2^j)^{-\tau} \leq \tfrac{1}{2}\gamma|k|^{-\tau}.$$

Therefore

$$|\langle \omega, k\rangle| \geq \gamma|k|^{-\tau} - \gamma\tfrac{1}{2}|k|^{-\tau} = \tfrac{1}{2}\gamma|k|^{-\tau}.$$

□

As lemma 2.4.4 sets a lower bound for the small denominators in the relevant domain for all truncations u^j, and f_k was already bounded from the Paley-Wiener lemma, all u^j are well-defined holomorphic functions.

With the above definitions of u^j and W_j, the only condition that remains to be proven for the application of the inverse approximation lemma is the geometric decay of the point-wise difference between u^j and u^{j-1}.

Ergo, the last thing we need to prove is that for every $\alpha \in \mathbb{R}_+ \setminus \mathbb{Z}_+$, there exists an N_α such that, for every $(x, \omega) \in (T^n) \times (\Gamma_\gamma + r_j)$,

$$\left|u^j(x, \omega) - u^{j-1}(x, \omega)\right| \leq N_\alpha r_j^\alpha. \tag{2.55}$$

For the version of Whitney inverse approximation lemma used in this thesis α cannot be an integer, see lemma 2.3.3.

The method of constructing u_j is very reminiscent of the method used to construct the approximate conjugation in Arnol'd's scheme. The frequencies are restricted to the shrinking domains $\Gamma_\gamma + r_j$ on which the frequencies satisfy part of the Diophantine conditions for lower $|k|$. This allows us to estimate $|u^j - u^{j-1}|$ by means of a Fourier series cut-off at $|k|$.

Estimates

In this paragraph, we prove super-exponential decay, which is a stronger condition than stated in (2.55), and sufficient to apply the inverse approximation lemma. We start by fixing j at some sufficiently high value and considering

$$\left|u^j(x,\omega) - u^{j-1}(x,\omega)\right| = \left|\sum_{2^{j-1}<|k|\leq 2^j} \frac{f_k(\omega)}{i\langle\omega,k\rangle} e^{i\langle x,k\rangle}\right| \leq \sum_{|k|>2^{j-1}} \left|\frac{f_k(\omega)}{i\langle\omega,k\rangle}\right| \cdot \left|e^{i\langle x,k\rangle}\right|. \tag{2.56}$$

For convenience, we included the rest of the series up to infinity in the last term. Substitute $|f_k|$ and $|i\langle\omega,k\rangle|$ by the estimates made in lemmas 2.4.3 and 2.4.4 to obtain

$$\sum_{|k|>2^{j-1}} \left|\frac{f_k(\omega)}{i\langle\omega,k\rangle}\right| \cdot \left|e^{i\langle x,k\rangle}\right| \leq \frac{2M}{\gamma} \sum_{|k|>2^{j-1}} |k|^\tau \cdot e^{(|\mathrm{Im}(x)|-\kappa)\cdot|k|} \leq \frac{2M}{\gamma} \sum_{|k|>2^{j-1}} |k|^\tau \cdot e^{-\frac{1}{2}\kappa\cdot|k|}, \tag{2.57}$$

where we used $\left|e^{i\langle x,k\rangle}\right| \leq e^{|\mathrm{Im}(x)|\cdot|k|} \leq e^{\frac{1}{2}\kappa\cdot|k|}$. We substitute $|k|$ by l, by using $\#\{k\in\mathbb{Z}^n : |k| = l\} \leq 2^n l^{n-1}$ as was done before in the proof of lemma 2.2.5, and we obtain

$$\sum_{|k|>2^{j-1}} |k|^\tau \cdot e^{-\frac{1}{2}\kappa\cdot|k|} \leq 2^n \sum_{l>2^{j-1}} l^{\tau+n-1} e^{-\frac{1}{2}\kappa l}. \tag{2.58}$$

We set $m = 2^{j-1}$ and as j is sufficiently high, we can state

$$\sum_{l=m+1}^{\infty} l^{\tau+n-1} e^{-\frac{1}{2}\kappa l} \leq \int_m^\infty s^{\tau+n-1} e^{-\frac{1}{2}\kappa s}\, ds = m^{\tau+n-1} e^{-\frac{1}{2}\kappa m} \int_m^\infty \left(\frac{s}{m}\right)^{\tau+n-1} e^{-\frac{1}{2}\kappa(s-m)}\, ds. \tag{2.59}$$

We use $\ln\left(\frac{s}{m}\right) \leq \frac{s}{m} - 1 = \frac{s-m}{m}$ and $\frac{\tau+n}{m} - \frac{1}{2}\kappa \leq 0$ for sufficiently high j to conclude that

$$\int_m^\infty \left(\frac{s}{m}\right)^{\tau+n-1} e^{-\frac{1}{2}\kappa(s-m)}\, ds = \int_m^\infty \exp\left(\ln\left(\tfrac{s}{m}\right)(\tau+n-1) - \tfrac{1}{2}\kappa(s-m)\right)\, ds \tag{2.60}$$

$$\leq \int_m^\infty \exp\left(\left(-\tfrac{1}{m} + \tfrac{\tau+n}{m} - \tfrac{1}{2}\kappa\right)(s-m)\right)\, ds \tag{2.61}$$

$$\leq \int_m^\infty \exp\left(-\tfrac{1}{m}(s-m)\right)\, ds = m. \tag{2.62}$$

Combining all of this (2.56, 2.57, 2.58, 2.59 and 2.62) results in

$$\left|u^j(x,\omega) - u^{j-1}(x,\omega)\right| \leq \frac{M2^{n+1}}{\gamma} m^{\tau+n} e^{-\frac{1}{2}\kappa m}, \tag{2.63}$$

which proves super-exponential decay for sufficiently high j.

Finishing the proof

Since $\left|u^j(x,\omega) - u^{j-1}(x,\omega)\right|$ decay super-exponentially for sufficiently high j, and super-exponential decay is much stronger than geometric decay, there exists a finite $J\in\mathbb{N}$ such that for all $j \geq J$ we have

$$\left|u^j(x,\omega) - u^{j-1}(x,\omega)\right| \leq N_\alpha r_j^\alpha. \tag{2.64}$$

Subsequently, we can pick each N_α large enough that the condition also holds for all $j < J$, as over a finite sequence the supremum is easily calculable. This proves the condition given by (2.55) for an arbitrary high $\alpha \in \mathbb{R}_+\backslash\mathbb{Z}_+$ and, in extension, that $\{u_j\}_{j\geq 0}$ indeed converges to a Whitney-smooth function $u(x,\omega)$ on $(T^n+\kappa)\times\Gamma_\gamma$. Lastly, we use the Whitney extension lemma to extend $u(x,\omega)$, which we have proven to exist and be Whitney-smooth on $T^n\times\Gamma_\gamma$, to a function on the entire space $T^n\times\Gamma$. This extension is not unique, and outside $T^n\times\Gamma_\gamma$, u does not satisfy (2.45). Finally, we set $\Gamma = \mathbb{R}^n$, which concludes the proof of theorem 2.4.2.

□ (of theorem 2.4.2)

Part II: Geometry

This part is dedicated to the geometric setting of KAM theory. In this thesis, KAM theory is introduced in terms of the flow of vector fields on manifolds, and this can introduce additional geometric considerations depending on the geometry of the manifold. This adds another significant part to KAM theory, in addition to the small denominator problem and the Whitney differentiability of the conjugation.

To begin, the vector fields that generate the flow of the dynamical system must be defined on a manifold M. Following the definition of continuous dimensional systems, the manifold M only needs to be locally diffeomorphic to a Banach space. However, in most KAM theorems and in this thesis, M is assumed to be a real, finite-dimensional, and connected manifold[1].

Additionally, M can be equipped with a structure $\mathcal{S}$. Not all structures can be defined on all manifolds, and only specific structures are proven to be compatible with KAM theory. Structures compatible with KAM theory are said to be admissible. The presence of an admissible structure on the manifold implies that only dynamical systems that preserve said structure are considered, i.e. only the vector fields for which the flow preserves the structure are considered.

KAM theory, like other perturbation theories, considers the perturbations of specific integrable vector fields. An integrable vector field must exhibit a quasi-periodic torus with parallel dynamics. In other words, there has to be a quasi-periodic torus proven to be present in the dynamical system before its persistence can be examined by KAM theory. The existence of such quasi-periodic tori is ensured by the definition of integrability in KAM theory. Furthermore, in the presence of a structure $\mathcal{S}$ on M, each integrable vector field must preserve $\mathcal{S}$.

KAM theorems are often stated for a specific manifold and structure. These manifolds and their associated structures are referred to as the context or setting of the KAM theorem. The context of a KAM theorem, or equivalently the geometry of the underlying manifold, also has a major impact on the external and internal parameters required within KAM theory. This has resulted in additional non-degeneracy conditions.

Outline of the part II

The first section introduces the four main contexts of KAM theory. These contexts vary and can be subdivided according to the dimension of M, the structure $\mathcal{S}$ associated with the manifold, and the specific geometry of the invariant torus.

In the second section, general Lie algebra methods for constructing structure-preserving diffeomorphisms are discussed, and the Lie homomorphism between functions and vector fields for Hamiltonian systems is briefly mentioned.

The third section discusses what is understood by an "integrable" system in KAM theory, and the associated Floquet form of such systems.

In the fourth section, parameters and frequency control are discussed. These lead to the non-degeneracy conditions required for traditional KAM theory.

The fifth section is about the Hamiltonian Lagrangian context and the resulting vector fields and Hamiltonian functions.

In the last section, an overview of the KAM method is given, describing the tactics used to obtain the desired conjugacy using the iterative KAM method. This section also discusses the differences between Kolmogorov's and Arnol'd's iteration schemes for the approximate conjugacies, followed by an overview of the Arnol'd scheme.

[1]KAM theory does exist for infinite-dimensional continuous dynamical systems. Discrete-time KAM theory exists and is the aforementioned KAM theory for maps.

3.1 Contexts of KAM theory

As mentioned before, KAM theory was originally developed in Hamiltonian systems. Since then, KAM theory has been generalized to other settings, such as the dissipative and volume-preserving setting. The setting of a KAM theorem is commonly referred to as the *context* of the KAM theorem, and this thesis follows that convention.

This section aims to explain how a context is defined within KAM theory and to briefly discuss the four main contexts of KAM theory.

3.1.1 Preliminaries

We begin with the basic setting and universal notation. A context of a KAM theorem consists the base manifold M, the structure $\mathcal{S}$ equipped on M, and the number of external[2] parameters involved.

The manifold

Let M be a real connected $(n+m)$-dimensional manifold containing a torus T^n, where we assume that $n \geq 2$ to allow for the existence of quasi-periodic dynamics on T^n. We equip this manifold with local coordinates $(x_1, \cdots, x_n, y_1, \cdots, y_m)$ such that the torus T^n is defined by $T^n = \{(x, y) \in M | y = 0\}$. This implies that $x \in T^n$, and thus that $x \pmod{2\pi} = x$, and $y \in \mathbb{R}^m$. Consequently, we can represent the manifold M as

$$M = T^n \times \mathbb{R}^m. \tag{3.1}$$

The manifold M can have a wide range of dimensions for both n and m. It is important to note that the choice of dimension has significant implications for the structures that can be equipped on M.

The structure

Additionally, we equip this manifold with a structure $\mathcal{S}$. The definition of a structure itself is quite loose, and many types of structures could be defined on M.

Definition 3.1.1 (Structure)**.** Let M be as described above. Then a structure $\mathcal{S}$ on M is an operator that maps one or more vectors in the tangent space of M at the point u, given by $T_u M$, to a value, which could be a real number, a vector field, or any other specified value.

For the purposes of KAM theory, the structure on M cannot be arbitrary and must satisfy certain conditions[3]. These conditions ensure that a KAM theorem can be proven for the context, and any structure that satisfies them is referred to as an *admissible* structure.

Definition 3.1.2 (Admissible structure)**.** Let $\mathcal{S}$ be a structure on M. Then $\mathcal{S}$ is an *admissible* structure if the following conditions hold:

- The $\mathcal{S}$-preserving vector fields on M form a Lie algebra $\mathfrak{g}$, which is a subalgebra of the Lie algebra of general vector fields on M. Associated with this Lie algebra is the closed Lie group of $\mathcal{S}$-preserving diffeomorphisms on M.
- The Fourier truncation and linearization of all vector fields in $\mathfrak{g}$ are contained within $\mathfrak{g}$.
- The integrable vector fields in $\mathfrak{g}$ can be written in the Floquet form with Floquet matrix Ω, and these Floquet matrices form a linear Lie algebra $\mathfrak{h}$. together with the commutator bracket. The Lie group H corresponding to $\mathfrak{h}$ must be an algebraic subgroup of $GL(m, \mathbb{R})$.

For a more in-depth description of these requirements, consult the studies by Broer, Huitema, and Takens [16–18]. While the definition is quite advanced, it is sufficient for now to know that the structures (with exception to the reversible structure) considered in this thesis do in fact meet these requirements.

Parameters

The parameters are given by $\nu \in \mathbb{R}^s$. They are relatively easy to define since they are unrestricted and remain fixed for most operations in KAM theory. These are addressed in later sections (section 3.5) and are not discussed further in this section.

[2] See section 3.5 for more details.

[3] Exceptions to this rule exist, most notably the reversible structure, which is not admissible.

Structure-preserving diffeomorphisms and vector fields

KAM theory, in the contexts where M is equipped with a structure $\mathcal{S}$, requires that the integrable vector fields and their perturbations preserve the structure. Moreover, the desired conjugations constructed by KAM theory in such contexts and their approximations are also expected to preserve the structure. The definitions of structure-preserving vector fields and diffeomorphisms are stated below.

Let M be as in (3.1) and let $\mathcal{S}$ be a structure defined on it.

Definition 3.1.3 (Structure-preserving diffeomorphism)**.** Let $f : M \to M$ be a diffeomorphism. If $\mathcal{S}$ is invariant under the push-forward of f, i.e. $f_*\mathcal{S} = \mathcal{S}$, then f is a *structure-preserving* diffeomorphism.

In the dissipative context where no structure is specified, we have that any diffeomorphism can be considered "structure-perserving".

Definition 3.1.4 (Structure-preserving vector field)**.** Let X be a vector field on M. If the flow φ_X^t of X on M for all t is a structure-preserving diffeomorphism, then X is a *structure-preserving* vector field.

Again, in the dissipative context, we have that any general vector field is "structure-perserving".

The four main contexts

The commonly encountered contexts in KAM theory can be classified into four main geometric settings. Although it is possible to define more exotic contexts that do not fit into these categories, they are rarely found in KAM theory literature[4]. These four general geometric settings are further subdivided into specific contexts based on the dimensions of M, the geometry of T^n, and the number of parameters.

Some of these contexts have been active areas of research until recently, for example, the reversible context 2 [88–90]. Others, such as the classical isotropic Hamiltonian context, are quite well understood, and modern research focuses more on the degenerate cases.

The four main geometric settings or contexts of KAM theory are:

- The dissipative context: This context is the simplest geometrically speaking, as in this context there is no structure present on the manifold M.
- The Hamiltonian context: This is the most well-known context for KAM theory and quasi-periodic motions in general. Due to Hamiltonian mechanics' close ties with classical physics and astronomy, Hamiltonian systems are among the most studied dynamical systems, and this pattern extends to KAM theory.
- The volume-preserving context: Volume-preserving vector fields are primarily found in the study of flows of incompressible fluids. Volume-preserving systems share considerable similarities with Hamiltonian systems, especially in the $m = 1$ case.
- The reversible context: Reversible systems display many of the characteristics of the other conservative systems. This similarity is profound enough that it is possible to prove KAM theorems within this context.

Each context can be and often is further subdivided based on the precise geometry of the manifold M and the invariant torus T^n.

3.1.2 Dissipative context

This context is characterized by the absence of a structure, which means that no structure-related restrictions are enforced on the vector fields and diffeomorphisms. As a result, KAM proofs in the dissipative context are usually simpler geometrically. KAM theory in this context was first developed by Moser [60, 61] and later by Broer, Huitema, and Takens [16, 18].

Parameters

The main difficulty in the dissipative context lies in the absence of suitable internal parameters, as for general integrable vector fields, the invariant quasi-periodic tori are isolated in phase space. This is quite different from the Hamiltonian context, where many quasi-periodic invariant tori can exist at the same time, parameterized by the action variable y.

To resolve the absence of internal parameters in the dissipative context, additional external parameters must be included in the context. These external parameters are then used to control the frequency of the quasi-periodic torus, allowing for the existence of a suitable family of quasi-periodic tori that range over the frequency space.

[4] Recently, KAM theory has been translated to a more exotic quantum context [102].

3.1.3 Hamiltonian context

The Hamiltonian context is defined by the preservation of the symplectic form.

Definition 3.1.5 (Symplectic form)**.** Let M be a $2N$-dimensional manifold and let ω^2 be a 2-from defined on M such that ω^2 is non-degenerate and ω^2 is closed.

A vector field X is said to be *symplectic* if the flow of X preserves ω^2.

The symplectic form induces a canonical coordinate system on the manifold M with well-known angle and action coordinates. Hamiltonian manifolds have N degrees of freedom when they have N pairs of action and angle coordinates, requiring every symplectic manifold to be $2N$-dimensional. In the Hamiltonian context, the manifold M can still be defined by (3.1), with the additional restriction that for $n \geq 2$ we have $m = n + 2p$ with $p \geq 0$.

Subdivisions

The Hamiltonian context is commonly subdivided based on the classification of T^n as a submanifold. In symplectic geometry, submanifolds are classified by their skew-orthogonal complement.

Skew-orthogonality and the skew-orthogonal complement are defined below.

Definition 3.1.6 (Skew-orthogonality)**.** Let $u \in M$ and $\xi, \eta \in T_uM$. If $\omega^2(\xi, \eta) = 0$, then ξ and η are *skew-orthogonal*.

Definition 3.1.7 (Skew-orthogonal compliment)**.** Let L be a submanifold of M and let T_uL be its tangent space at the point $u \in L$. The set of all tangent vectors $\xi \in T_uM$ such that ξ is skew-orthogonal to each $\eta \in T_uL$ is called the *skew-orthogonal complement* of T_uL, denoted by $(T_uL)^\perp$.

Again, let L be a submanifold of M and let T_uL be its tangent space at the point $u \in L$. We classify L in the following way:

- Isotropic submanifold: If all tangent vectors in T_uL are contained in the skew-orthogonal complement $(T_uL)^\perp$ for all points $u \in L$, then L is an *isotropic* manifold. This also implies that the dimension of L is lower than or equal to N.
- Coisotropic submanifold: If the skew-orthogonal complement $(T_uL)^\perp$ is contained in T_uL for all points $u \in L$, then L is a *coisotropic* manifold. The dimension of a coisotropic manifold is at least N.
- Lagrangean submanifold: If the skew-orthogonal complement $(T_uL)^\perp$ is equal to T_uL for all points $u \in L$, then the dimension of L is equal to N and L is a *Lagrangian* manifold. Lagrangean submanifold are also referred to as maximal isotropic submanifolds.
- Antropic submanifold: Submanifolds which are neither isotropic or coisotropic. These are unfortunately not further discussed in this thesis.

KAM theorems have been proven for both isotropic and coisotropic invariant tori. These differ significantly and are thus treated as separate contexts within KAM theory.

The isotropic Hamiltonian context is the most well-understood. The symplectic form has a universal normal form and can be assumed to be independent of any parameters. By Herman's lemma[5], any non-resonant torus with parallel flow of a Hamiltonian system is isotropic provided that the symplectic form is exact. See [31, 46, 47, 87] for a proof and [19] for discussion. Most KAM theorems for the Hamiltonian context only consider the isotropic invariant tori, and examples are widely available; see [17].

The coisotropic Hamiltonian context is more complex. The submanifolds are dependent on a specific symplectic form, and the behavior of Hamiltonian systems on coisotropic manifolds is very sensitive to this symplectic form in regards to the arithmetic properties. For example, every vector in the submanifold could be resonant. Unlike the isotropic context, no universal symplectic form is available. For more information and KAM theory for coisotropic tori, see the articles by Herman [48, 49] and Parasyuk [66–68].

Lagrangian tori[6] are treated the same as isotropic tori, so these also fall under the isotropic context. KAM theory in the Lagrangian context, where the invariant torus is a Lagrangian submanifold, is usually referred to as classical KAM theory, as this context was the first for which KAM theory was proven by Kolmogorov and Arnol'd. Examples for KAM theory in this context are even more accessible, and most tutorials [19, 56, 73] on KAM theory start with this context.

For more information on the Hamiltonian contexts in general, see the book [17] or, for a more recent overview, see the study [91] by Sevryuk.

[5] This lemma is also known as the automatic isotropicity lemma.

[6] Lagrangian tori are also referred to as *maximal* tori, due to them being maximal isotropic submanifolds.

3.1.4 Volume-preserving context

The volume preserving context is defined by the preservation of the volume form.

Definition 3.1.8 (Volume form)**.** Let M be an N-dimensional manifold and let ω^N be an N-form defined on M such that ω^N is non-degenerate[7].

A vector field X is said to be *volume-preserving* if the flow of X preserves the volume form ω^N. This type of vector fields is common in the study of incompressible fluids. In the volume-preserving context, the manifold M is defined as in (3.1), with the additional restriction that $m \geq 1$ and the usual restriction that $n \geq 2$.

Subdivisions

Within this context, there are two distinct cases depending on the codimension of the torus. Let M be as defined in (3.1), and subject to the additional restriction.

- The $m = 1$ case: In this case, the codimension of the torus T^n is one. This case is closely related to the Hamiltonian context, as the volume form on $T^n \times \mathbb{R}$ behaves very similarly to the Hamiltonian case.
- The $m > 1$ case: In this case, the codimension of the torus T^n is greater than one. This context is treated similarly to the general dissipative case within KAM theory.

Research

KAM theory within the volume-preserving context has been studied by Moser [60, 61] and this research was continued by Broer and Braaksma [9, 12, 13].

3.1.5 Reversible context

The reversible context is defined by the preservation of the following structure.

Definition 3.1.9 (Reversible structure)**.** Let $M = T^n \times \mathbb{R}^{u+v}$ with $u, v > 0$ and let G be some fixed involution of the type $(n + v, u)$.

A vector field X is *G-reversible* if G transforms X into the opposite field $-X$, or equivalently if

$$(D_u G)X(a) \equiv -X(G(a)) \quad \text{for all} \quad a \in M. \tag{3.2}$$

Technically, any diffeomorphism $G : M \to M$ can be used within the definition, however usually only involutions, $G^2 = \mathrm{id}$, are considered. Additionally, it is assumed that the invariant n-tori with parallel flow are invariant under both the regular flow and reversed flow of X.

Far more importantly, the G-reversible vector fields do *not* form a Lie algebra. This complicates the construction of structure-preserving transformations, which must then be carried out in the Lie algebra of general vector fields. For more details consult the study [16] by Broer and Huitema, in which the Lie algebra method is adapted to reversible systems.

Subdivision

The reversible context can be subdivided into two distinct cases, depending on the size and shape of the matrix G and the number of external parameters available. As these contexts are extensive and difficult to fully explore, they are not covered in this thesis.

- Reversible context 1: This context is defined by $u \geq v$.
- Reversible context 2: This context is defined by $s \geq v - u > 0$, where s is the number of external parameters. Note that this context requires at least $v - u$ external parameters.

If $v - u > s > 0$, i.e. there are not enough external parameters available, then the G-reversible vector fields have no invariant tori and this case is therefore not studied in KAM theory. For details, on these cases and the reversible context in general consult the book [17] by Broer, Huitema and Sevryuk.

Proving KAM theorems in the reversible context with the Lie algebra method requires additional arguments, as the reversible vector field do not naturally form a Lie algebra. This is beyond the scope of this thesis and therefore the reversible context is not further mentioned. Lastly, for those interested in these arguments, please refer to [16].

Research

The reversible context is the most studied context for KAM theory after the Hamiltonian context. Again, Moser [60, 61] was the first to study this context, followed by Bibikov and Pliss [6, 7]. Later, this research was continued by Scheurle [82–84], Parasyuk [65], Pöschel [71], Arnol'd and Sevryuk [3, 4, 86], among others.

[7] On N-dimensional manifolds the volume form ω^N naturally satisfies $d\omega^N = 0$ and is therefore closed.

3.2 More on symplectic geometry

This section introduces the general definitions and methods from symplectic geometry used by KAM theory for the Hamiltonian context.

To begin, we formally define the symplectic manifold and its structure. In KAM theory for the Hamiltonian context, the canonical symplectic form is used. This form is widely used due to its additional properties, most notably being an exact differential form. This property allows for the use of the generating function method for constructing (exact) symplectomorphisms. The definition of symplectomorphisms is also given.

The discussion then moves to general vector fields and their characterizations, where the definition of symplectic vector fields is given. Finally, Hamiltonian vector fields and diffeomorphisms are defined and discussed, representing a further refinement of general symplectic vector fields and symplectomorphisms.

3.2.1 Preliminaries

Let M be a connected $2N$-dimensional real manifold.

Definition 3.2.1 (Symplectic structure)**.** A symplectic structure on M is defined as a closed, non-degenerate, differentiable two-form ω^2 on M. A manifold equipped with a symplectic structure is referred to as a *symplectic* manifold.

Canonical symplectic form

Not all symplectic forms are equally important, some symplectic forms are closely associated with physics and therefore more common and well-studied. Symplectic geometry has its origins in classical mechanics, particularly Hamiltonian mechanics, and associated with these fields is the canonical symplectic form.

Definition 3.2.2 (Canonical symplectic form)**.** Let M be as above, then the canonical symplectic form is given by

$$\omega^2 = \sum_{i=1}^{N} dx_i \wedge dy_i. \tag{3.3}$$

Moreover, the canonical symplectic form is commonly used because all symplectic forms can be locally reduced to it by means of Darboux's theorem.

Theorem 3.2.3 (Darboux)**.** *Given a symplectic form ω^2 and a point x_0, there exists a local diffeomorphism f on a neighborhood of x_0 in M such that the pull-back of ω^2 by f, denoted by $f^*\omega^2$, is of the form in* (3.3)*.*

Therefore, the canonical symplectic form is also used in the Hamiltonian context of KAM theory.

Symplectomorphisms

The canonical sympletic form has the following property.

Definition 3.2.4 (Exact differential form)**.** Let ω be a differential n-form, ω is *exact* if it is the exterior derivative of another differential form, i.e. $\omega = d\alpha$, where α is a differential $(n-1)$-form. A manifold equipped with an exact n-form is referred to as an *exact manifold*.

It is easy to verify that the canonical symplectic form is exact with $\alpha = -\sum_i y_i dx_i$. With this property, we can define (exact) symplectomorphisms.

Definition 3.2.5 (Symplectomorphism)**.** Let M be as above, let $\omega^2 = d\alpha$ be an exact symplectic form, and let $f : M \to M$ be a diffeomorphism, then the following holds.

- f is a *symplectomorphism* if and only if $f^*\omega = \omega$ (and in extension the form $f^*\alpha - \alpha$ is closed).
- f is a *exact symplectomorphism* if and only if $f^*\theta - \theta = dS$, where S is a function and referred to as the *primitive* function of f[8].

The composition of two symplectomorphisms is a symplectomorphism, and symplectomorphisms on a given manifold are invertible. Therefore, they form a group denoted by $\mathrm{Symp}(M, \omega^2)$.

[8] This is not equivalent to a generating function.

3.2.2 Symplectic vector fields

Lemma 3.2.7 is usually proven for general differential forms, and this thesis follows this convention. Therefore, in this subsection, let M be an N-dimensional manifold equipped with a non-degenerate closed differential n-form ω with $2 \leq n \leq N$. This is explicitly stated in the definition and the lemma.

We say that a vector field X on M *preserves* the differential form if ω is invariant under the pull-back of the flow of X, i.e. $(\varphi_X^t)^*\omega = \omega$, where φ_X^t is the flow of X on M.

Identification of vector fields by a differential form.

A closed differential n-form ω allows us to identify a vector field on M with an $(n-1)$-form.

Definition 3.2.6 (Identification with form)**.** Let M be an N-dimensional manifold and ω a non-degenerate closed differential n-form such that $N \geq n \geq 2$. Lastly, let X be an arbitrary vector field on M. We define

$$\mathcal{I}_\omega(X) := \mathbf{i}_X\omega, \tag{3.4}$$

where $\mathbf{i}_X$ is the interior product (see definition B.4.1.).

Preserving a differential form

The identification allows us to quickly determine if a vector field X on M preserves the differential form ω.

Lemma 3.2.7. *Let M be an N-dimensional manifold, and let ω be a non-degenerate closed differential n-form such that $N \geq n \geq 2$. A vector field X on M preserves ω if and only if the differential $(n-1)$-form $\mathcal{I}_\omega(X)$ is closed.*

Proof. Let us denote the flow of X on M by φ_X^t. We use Cartan's magic formula for the Lie derivative (See appendix B.4.4)

$$\mathcal{L}_X\gamma = d(\mathbf{i}_X\gamma) + \mathbf{i}_X d\gamma,$$

and the closedness of ω, i.e. $d\omega = 0$, to state that

$$\frac{d}{ds}\left(\varphi_X^{t+s}\right)^*\omega\Big|_{s=0} = (\varphi_X^t)^*\mathcal{L}_X\omega = (\varphi_X^t)^*\left(d(\mathbf{i}_X\omega) + \mathbf{i}_X d\omega\right) = (\varphi_X^t)^* d\mathcal{I}_\omega X. \tag{3.5}$$

If $\mathcal{I}_\omega(X)$ is closed, then we have $d\mathcal{I}_\omega = 0$. This implies that $\frac{d}{ds}\left(\varphi_X^{t+s}\right)^*\omega\big|_{s=0} = 0$ and therefore that ω is constant under the flow. Thus, by definition, X preserves the differential form.

If X preserves the differential form, then we have $(\varphi_X^t)^*\omega = \omega$. This implies that $\frac{d}{ds}\left(\varphi_X^{t+s}\right)^*\omega\big|_{s=0} = 0$ and in extension $(\varphi_X^t)^* d\mathcal{I}_\omega X = 0$. Therefore we have $d\mathcal{I}_\omega = 0$ meaning that $\mathcal{I}_\omega(X)$ is closed. □

We say that a vector field X is *symplectic* if it preserves the symplectic form ω^2. Similarly, we say that a vector field X is *volume-preserving* if it preserves the volume form ω^N.

3.2.3 Hamiltonian vector fields and diffeomorphisms

In this subsection, let M be a symplectic manifold equipped with a symplectic form ω^2.

Hamiltonian vector fields

Definition 3.2.8 (Hamiltonian vector field)**.** Let X be a vector field on M. If there exists a function H such that

$$\mathcal{I}_{\omega^2}X = -dH, \tag{3.6}$$

then X is a *Hamiltonian* vector field[9] and H is referred to as the *Hamiltonian* or *Hamiltonian function* of X.

If H does not exist, but $\mathcal{I}\omega^2 X$ is closed, then X is said to be *locally Hamiltonian*. As the name implies, locally Hamiltonian vector fields can be expressed locally by a Hamiltonian function, and all symplectic vector fields are locally Hamiltonian by lemma 3.2.7.

Suppose X is a Hamiltonian vector field and ω^2 is the canonical symplectic form (3.3). Then H is the familiar Hamiltonian function, and X can be stated by the familiar representation of a Hamiltonian vector field:

$$X_{x_i} = \frac{\partial H}{\partial x_i} \quad X_{y_i} = -\frac{\partial H}{\partial y_i}. \tag{3.7}$$

[9] Or equivalently, an exact symplectic vector field

Another useful property of Hamiltonian vector fields is that the push-forward of the vector field by any symplectomorphism remains Hamiltonian.

Suppose f is a symplectomorphism, i.e $f_*\omega = \omega$, and X is a Hamiltonian vector field with $\mathcal{I}_\omega(X) = \mathbf{i}_X\omega = -dH$. Then, the push-forward of X by f satisfies

$$\mathbf{i}_{f_*X}\omega = \mathbf{i}_{f_*X} f_*\omega = f_*(\mathbf{i}_X\omega) = -f_*dH = -df_*H = -d(H \circ f^{-1}). \tag{3.8}$$

Therefore, f_*X is a Hamiltonian vector field with the Hamiltonian $H \circ f^{-1} = (f^{-1})^*H$. This allows us to do most calculations on a Hamiltonian vector field by applying symplectic coordinate transformations on the Hamiltonian.

Hamiltonian diffeomorphisms

Closely associated with the Hamiltonian vector fields are the Hamiltonian diffeomorphisms.

Definition 3.2.9 (Hamiltonian diffeomorphism)**.** Let X_H be a Hamiltonian vector field with the Hamiltonian function H, and let $t \in \mathbb{R}$. The flow of X_H, denoted by φ_H^t, over time t is referred to as a *Hamiltonian diffeomorphism* corresponding[10] to the Hamiltonian H.

Hamiltonian vector field and diffeomorphisms preserve the symplectic form.

Lemma 3.2.10. *Suppose X_H is a Hamiltonian vector field with the Hamiltonian function H, and φ_H^t are the associated Hamiltonian diffeomorphisms. Then, X_H preserves the symplectic form, and all associated Hamiltonian diffeomorphisms are symplectomorphisms.*

Proof. It is easy to verify that Hamiltonian vector fields preserve the symplectic form since all functions, or equivalently 0-forms, have the property that $d(df) = 0$. Therefore, as $\mathcal{I}\omega^2 X_H = -dH_X$, we have $d\mathcal{I}\omega^2 X_H = -d(dH_X) = 0$. We can easily substitute this in the proof of lemma 3.2.7 to determine that the vector field X_H indeed preserves the symplectic form.

Lastly, by definition of the symplectic vector field, we have that the Hamiltonian diffeomorphisms φ_H^t are symplectomorphisms. □

This lemma allows us to quickly generate symplectomorphism by considering the Hamiltonian diffeomorphisms with $t = 1$, i.e. the time-one map of Hamiltonian vector fields. This is the most common method of generating symplectomorphisms or other structure-preserving diffeomorphism, and the method is also used in this thesis.

Lie algebra and Lie group

The last point discussed is that the Hamiltonian vector fields form a subalgebra of the Lie algebra of general vector fields, when combined with the Lie bracket for vector fields.

Additionally, the Hamiltonian diffeomorphisms form a closed subgroup of $Symp_0(M, \omega^2)$ and, more importantly, are the Lie group associated with the Hamiltonian vector field Lie algebra. These statements are given here without proof but can easily be verified. See the study by Broer, Huitema, and Takens [18] for details.

For more theory on symplectic manifolds and Hamiltonian systems in general, see the book [57] by Marsden and Ratiu.

3.2.4 Remarks

Volume-preserving case.

The volume-preserving context is very similar to the Hamiltonian context. Therefore, most of the theory can be easily generalized to generic n-forms, including the volume form, as demonstrated in Lemma 3.2.7.

Generating function method for symplectomorphisms.

The time-one flow map method is not the only way to construct symplectomorphisms. An older method of constructing symplectomorphisms is the generating function method. This method generates exact symplectomorphisms by solving $f^*\theta - \theta = dS$.

The generating function method has fallen out of favor compared to the time-one flow map method. However, it was widely used in earlier publications on KAM theory. For example, the original papers by Kolmogorov, Arnold, and Moser are all formulated with the generating function method to construct the symplectomorphisms.

Since this method is not used in this thesis, it is not further mentioned.

[10] There are infinitely many Hamiltonian diffeomorphisms associated with a Hamiltonian function, one for each possible time value.

3.3 Lie algebras of vector fields

To easily generate structure-preserving diffeomorphisms, it is convenient to consider the vector fields on the N-dimensional manifold M equipped with the structure $\mathcal{S}$ as a Lie algebra. This has several advantages.

The foremost advantage is that it generalizes the method used to construct structure-preserving diffeomorphisms for all four main contexts, which allows for a generalized KAM method.

Secondly, it allows us to simplify the construction of structure-preserving diffeomorphisms when compared to older methods. A particular example of this is the time-one flow map method in the Hamiltonian context, which is widely applied in modern KAM literature.

3.3.1 Lie algebra

Recall that a Lie algebra is a vector space $\mathfrak{g}$ together with a Lie bracket, a binary operator $[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, which is bi-linear, alternating, and satisfying the Jacobi identity.

It is easy to verify that general vector fields over a manifold M form a Lie algebra with the following operator:

Definition 3.3.1 (Lie bracket for vector fields)**.** Let X, Y be vector fields over M. We define

$$[X,Y](f) = X(Y(f)) - Y(X(f)) \quad \text{for all} \quad f \in C^\infty(M), \tag{3.9}$$

or equivalently point-wise for $x \in M$,

$$[X,Y]_x = (\mathcal{L}_X Y)_x = \left.\frac{d}{dt}\right|_{t=0} \left(\left(D\varphi_X^t\right)_{\varphi_X^{-t}(x)} Y_{\varphi_X^{-t}(x)} \right), \tag{3.10}$$

where $\mathcal{L}_X$ is the Lie derivative at point x and $\left(D\varphi_X^t\right)_{\varphi_X^{-t}(x)}$ is the total derivative at $\varphi_X^{-t}(x)$. The operator $D\varphi_X^t$ can also be referred to as the push-forward of φ_X^t, denoted by $\left(\varphi_X^t\right)_*$.

The subset of vector fields over M that preserve an admissible structure $\mathcal{S}$ form a Lie subalgebra of the Lie algebra of general vector fields. It is known that the Hamiltonian and globally divergence-free vector fields form subalgebras of the Lie algebra of general vector fields. However, the reversible vector fields do not form a Lie algebra and require more complicated methods to apply the Lie algebra method. For a complete description of the Lie algebra used in this case, see [16].

3.3.2 Lie group

Associated with the Lie algebra of vector fields is the Lie group of the diffeomorphisms defined by the flows of the vector fields.

Definition 3.3.2 (Flow mapping)**.** Let $t \in \mathbb{R}$.

$$\varphi^t : \mathfrak{g} \to G : X \mapsto \varphi_X^t, \tag{3.11}$$

where φ_X^t is the flow of the vector field X on M.

The Lie algebra of vector fields is the infinitesimal generator of the Lie group, which means that the derivative of a diffeomorphism contained within the Lie group with respect to time t results in the vector field. Thus, for a vector field $X \in \mathfrak{g}$ and the flow of X on M at time t, denoted by φ_X^t, we have the identity

$$\left(d\varphi_X^t(x)\right)_{t=0} = \left.\frac{\partial}{\partial t}\varphi_X^t(x)\right|_{t=0} = X(x) \quad \text{for all} \quad x \in M. \tag{3.12}$$

3.3.3 Exponential map

It is possible to define the exponential map for the Lie algebra of vector fields and the corresponding Lie group of diffeomorphisms. In KAM theory and most other applications of Lie algebras of vector fields, it is common to define the exponential map as the time-one flow map of the vector field

$$\exp : \mathfrak{g} \to G : X \mapsto \exp X = \varphi_X^1. \tag{3.13}$$

The exponential map is a map between the Lie algebra of vector fields and the Lie group of diffeomorphism defined by their flow, which preserves the local group structure from the Lie algebra. This definition of the exponential map is also applicable to the subalgebras associated with admissible structures.

3.3.4 Adjoint representation

The last property of Lie algebras that is discussed is the adjoint representation, which can be used to encode conjugacies of vector fields.

Group action

We first consider the group action of the Lie group of diffeomorphisms on itself by conjugation.

Definition 3.3.3 (Group action)**.** Consider $\Psi : G \to \mathrm{Aut}(G)$, which maps elements of G to automorphisms of G. We define this map by $\Psi : g \mapsto \Psi_g$ such that for all $g, h \in G$ the automorphism Ψ_g is given by the conjugacy of h by g, i.e.

$$\Psi : G \to \mathrm{Aut}(G) : g \mapsto \Psi_g \quad \text{where} \quad \Psi_g(h) := ghg^{-1}. \tag{3.14}$$

It is possible to construct other group actions of G on itself. However, the adjoint representation is only defined for this particular group action. Fortunately, we are only interested in conjugations within KAM theory.

Adjoint representation of the Lie group

With the group action, we can determine the adjoint action, which is the action of the Lie group on the Lie algebra.

Definition 3.3.4 (Adjoint representation)**.** Let $t \in \mathbb{R}$ and let Ψ be as defined in (3.14). Then, the adjoint action of the Lie group, denoted by Ad, is given by

$$\mathrm{Ad} : \begin{array}{ccc} G & \to & \mathrm{Aut}(\mathfrak{g}) \\ g & \mapsto & \mathrm{Ad}_g \end{array} \quad \text{such that} \quad \begin{array}{ccc} \mathfrak{g} & \xrightarrow{Ad_g} & \mathfrak{g} \\ {\scriptstyle \varphi^t}\downarrow & & \downarrow{\scriptstyle \varphi^t} \\ G & \xrightarrow{\Psi_g} & G \end{array} \quad \text{commutes.} \tag{3.15}$$

This definition can be examined point-wise to determine $\mathrm{Ad}_g(X)$ explicitly. We have that

$$\begin{aligned} \left.\frac{\partial}{\partial t}\right|_{t=0} \Psi_g\left[(\varphi_X^t)(x)\right] &= \left.\frac{\partial}{\partial t}\right|_{t=0} \left(g \circ \varphi_X^t \circ g^{-1}(x)\right) && (3.16)\\ &= (Dg)_{\varphi_X^0 \circ g^{-1}(x)} \left.\frac{\partial}{\partial t}\right|_{t=0} \left(\varphi_X^t \circ g^{-1}(x)\right) && (3.17)\\ &= (Dg)_{g^{-1}(x)} X_{g^{-1}(x)} && (3.18)\\ &= \mathrm{Ad}_g(X)(x), && (3.19) \end{aligned}$$

where we have used the chain rule and (3.12). Therefore, $\mathrm{Ad}_g(X) = (Dg)_{g^{-1}(x)} X_{g^{-1}(x)}$.

Adjoint representation of the Lie algebra

The adjoint action of the Lie group G allows us to define the adjoint representation of the Lie algebra $\mathfrak{g}$ by means of the identity (3.12).

Definition 3.3.5 (Adjoint action)**.** Let the adjoint representation Ad be as defined in (3.15). Then, the adjoint action of the Lie algebra, denoted by ad, is given by

$$\mathrm{ad} : \begin{array}{ccc} \mathfrak{g} & \to & \mathrm{Aut}(\mathfrak{g}) \\ X & \mapsto & \mathrm{ad}_X \end{array} \quad \text{such that} \quad \begin{array}{ccc} \mathrm{ad}_X : & \mathfrak{g} & \to & \mathfrak{g} \\ & \mathrm{ad}_X & := & \left.\frac{\partial}{\partial t}\right|_{t=0} \mathrm{Ad}_{\varphi_X^t} \end{array} \quad \text{holds.} \tag{3.20}$$

In our case, this definition results in

$$\mathrm{ad}_X(Y)(x) = \left.\frac{\partial}{\partial t}\right|_{t=0} \mathrm{Ad}_{\varphi_X^t}(Y)(x) = \left.\frac{\partial}{\partial t}\right|_{t=0} \left(\left(D\varphi_X^t\right)_{\varphi_X^{-t}(x)} Y_{\varphi_X^{-t}(x)}\right) = [X, Y]_x. \tag{3.21}$$

Therefore, $\mathrm{ad}_X(Y) = [X, Y]$.

3.3.5 The conjugation property in Lie algebra terms

The adjoint representation of the Lie group allows us to describe conjugation between vector fields in terms of the Lie algebra. This generalizes the calculation and methods used to calculate the conjugation.

Consider the conjugation of the flows of vector fields X and $\tilde{X}$, by the push-forward of a diffeomorphism g:

$$(g_*)_x X_x = g_* X_x = \tilde{X}_{g(x)}, \tag{3.22}$$

which is the general conjugation property for vector fields. In terms of the adjoint representation, (3.22) is equivalent to

$$\mathrm{Ad}_g(X) = \tilde{X}. \tag{3.23}$$

Calculation of small structure-preserving conjugations

Let $\tilde{X}$ be a small perturbation of X, and assume the diffeomorphism g is given by the time-one flow map of V, i.e. $g = \varphi_V^1$, where$V$ is a (structure-preserving) vector field in the Lie algebra.

We can express $\mathrm{Ad}g(X)$ entirely in terms of the vector field V. This is achieved by expanding $\mathrm{Ad}\varphi_V^t(X)$ about time $t = 0$ and evaluating it at time $t = 1$, resulting in

$$\mathrm{Ad}_{\varphi_V^1}(X) = X + \left.\frac{\partial}{\partial t}\right|_{t=0} \mathrm{Ad}_{\varphi_V^t}(X) + \int_0^1 (1-t) \left.\frac{\partial^2}{\partial t^2}\right|_{t=0} \left(\mathrm{Ad}_{\varphi_V^t}(X)\right)\, dt \tag{3.24}$$

$$= X + \mathrm{ad}_V(X) + \int_0^1 (1-t) \cdot \mathrm{ad}_V(\mathrm{ad}_V(X))(\varphi_V^t)\, dt \tag{3.25}$$

$$= X + [V, X] + \int_0^1 (1-t) \cdot [V, [V, X]]_{\varphi_V^t}\, dt. \tag{3.26}$$

It follows that we can determine the vector field V, and in extension the whole conjugation, by solving the following equation:

$$\mathrm{Ad}_{\varphi_V^1}(X) - \tilde{X} = \mathrm{ad}_V(X) - (\tilde{X} - X) + O(V^2) = 0. \tag{3.27}$$

Lastly, since the difference between X and $\tilde{X}$ is small, and in extension the diffeomorphism g and vector field V are small, we can approximate V to a sufficient degree by solving:

$$\mathrm{ad}_V(X) = \tilde{X} - X. \tag{3.28}$$

This is equation is referred to as the linearized conjugation property[11], and it is common place in many KAM theorems.

3.3.6 Lie homomorphism for the Hamiltonian vector fields

Suppose the Lie algebra and Lie group are the Lie algebra of Hamiltonian vector fields and the Lie group of Hamiltonian diffeomorphisms, respectively, then solving (3.28) becomes particularly simple.

This is because Hamiltonian vector fields are uniquely determined by their Hamiltonian function, which allows for the existence of a Lie algebra homomorphism $\mathcal{F} : \mathfrak{f} \to \mathfrak{g}$ between the Lie algebra $\mathfrak{f}$ of Hamiltonian functions with the (negative) Poisson bracket and the Lie algebra $\mathfrak{g}$ of Hamiltonian vector fields with the Lie bracket[12]. Here, the Poisson bracket is defined as follows.

Definition 3.3.6 (Poisson bracket)**.** For all $f, g \in \mathfrak{f}$, we define

$$\{f, g\} = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \frac{\partial g}{\partial y_i} - \frac{\partial g}{\partial y_i} \frac{\partial f}{\partial x_i} \right) \quad \text{and have} \quad [X_f, X_g] = X_{\{g,f\}}. \tag{3.29}$$

The Lie homomorphism $\mathcal{F}$ allows us to calculate all operations on the Lie algebra vector fields, such as solving (3.28), by performing the equivalent calculations on the Lie algebra of Hamiltonian functions, where the calculations are considerably easier.

For example, if $H, \tilde{H}, F$ are the Hamiltonian functions for the Hamiltonian vector fields $X, \tilde{X}, V$ respectively, then the conjugation property (3.27) is equivalent to

$$H \circ \exp(\bar{\Psi}) = \tilde{H} + \{H, F\} + \int_0^1 (1-t) \cdot \{\{H, F\}, F\} \circ \exp(t\bar{\Psi})\, dt. \tag{3.30}$$

[11] Within KAM theory, it is more commonly known as the homological equation.

[12] Note that only the bracket is inverted after applying the homomorphism. This can be quite confusing, and is direct result from using the traditional definitions for the Poisson bracket and the Lie bracket for vector fields.

3.4 Integrability within KAM theory

We have discussed the context of KAM theory and the associated geometric considerations. This section is about integrability of vector fields within KAM theory.

Integrability within KAM theory

As with many perturbation theories, KAM theory concerns itself with perturbations of the well-understood unperturbed or "integrable" system. The concept of integrability within KAM theory is specific to KAM theory, as it specifically defines the systems on which KAM theory is applicable.

KAM-integrability, which continuing forward is referred to as just integrability, implies that the vector field contains invariant quasi-periodic torus with parallel flow within their dynamics. This is more abstractly defined as follows.

Definition 3.4.1 (Integrability)**.** Let X be a vector field over M with an invariant torus T^n. If the flow of X on T^n is equivariant under the free action of T^n then X is *integrable* at T^n.

In structure-preserving contexts, we additionally require that all integrable vector fields X preserve the structure defined on M.

Parameterized KAM theory

Most modern KAM theorems are parameterized versions of KAM theory. Firstly, because parameterized variants of KAM are much more natural in many applications. Secondly, in some contexts, such as the dissipative context and the reversible context 2, additional external parameters are even required to prove a (meaningful) KAM theorem. Lastly, it is relatively straightforward to adapt parameterized KAM theorems to specific cases by mapping the internal variables to the parameters.

Therefore, the parameters $\mu \in P \subset \mathbb{R}^s$ are introduced, where P is assumed to contain μ_0.

Parameterized KAM theory considers integrable families of vector fields $\{X^\mu\}_\mu$ parameterized by μ, and determines for which values of μ the invariant torus survives under small perturbations. Integrable families must satisfy a few properties, which are given below.

Definition 3.4.2 (Integrable s-parameter family)**.** Let $\{X^\mu\}_\mu$ be an s-parameter family of vector fields. $\{X^\mu\}_\mu$ is an integrable family of vector fields if and only if each X^μ is integrable, and $X^{\mu_0} \in \{X^\mu\}_\mu$ induces parallel dynamics on the invariant torus[13].

In the usual local coordinates, see section 3.1.1, integrable s-parameter families of vector fields $\{X^\mu\}_\mu$ are explicitly given by

$$X^\mu = f(y,\mu)\frac{\partial}{\partial x} + g(y,\mu)\frac{\partial}{\partial y}, \tag{3.31}$$

where $g(0,0) = 0$ and $f(0,\mu) =: \omega(\mu) \neq 0$.

Vertical vector fields

An analytic s-parameter family of integrable vector fields is often regarded as a "vertical" vector field on $M \times P$ within KAM theory.

Definition 3.4.3 (Vertical vector field)**.** Suppose $\{X^\mu\}_\mu$ is an analytic s-parameter family of integrable vector fields on phase space M and parameter space P. Then

$$X : M \times P \to TM \times TP : ((x,y),\mu) \mapsto (X^\mu(x,y),(\mu,0)) \tag{3.32}$$

is referred to as an analytic *vertical* vector field.

The motivation behind this terminology comes from the circle homomorphism example, where the circle and parameter space together are interpreted as a vertical cylinder.

The invariant n-torus for each X^μ is usually denoted by V^μ, and these form a smooth family $V_\mu \times \mu$ of X^μ-invariant n-tori. The submanifold $V = \bigcup_{\mu\in P} V_\mu \times \{\mu\} \subset M \times P$ is invariant under the flow of the vertical vector field X.

Vertical vector fields and the associated invariant submanifolds are relatively common in KAM literature. However, in this thesis, we assume that the same torus T^n is X^μ-invariant for all $\mu \in P$. Therefore, there is no need to use the more complex vertical vector fields.

[13] It is not necessary to assume that each X^μ has the same invariant torus T^n. However, in this thesis, this assumption is made.

3.4.1 Floquet tori and normal frequencies

Closely related to integrability in KAM theory is Floquet reducibility.

Definition 3.4.4 (Floquet tori)**.** Let $M = T^n \times \mathbb{R}^m$ be equipped with the usual local coordinates, and let X be a vector field on M such that T^n is an invariant torus with parallel dynamics. The torus T^n is referred to as a *Floquet* torus if the vector field X determines a system of equations in the so-called *Floquet* form, which is given by

$$\dot{x} = \omega + O(y) \tag{3.33}$$

$$\dot{y} = \Omega y + O(y^2), \tag{3.34}$$

with $\Omega \in \mathrm{gl}(m, \mathbb{R})$ and where Ω is referred to as the *Floquet matrix* of the torus.

The Floquet form is defined by the independence of Ω from x. The vast majority of KAM theory only considers invariant quasi-periodic Floquet tori, and the definition of integrability reflects this. The invariance under the action of T^n implies that ω and $D_y g(0,0) = \Omega$ are x-independent.

Associated with the Floquet matrix are the normal frequencies.

Definition 3.4.5 (Normal frequencies)**.** The *normal* frequencies of a Floquet torus are the positive[14] imaginary parts of the eigenvalues of its Floquet matrix Ω.

These frequencies are found in KAM theory for lower dimensional tori, i.e. $m \geq 0$, where they can result in small denominators. This needs to be taken into account in the arithmetic conditions. Lower dimensional tori and normal frequencies are outside the scope of this thesis unfortunately.

The frequency of the torus itself, given by ω in (3.33), are usually referred to as the *internal* frequencies of the torus.

Dissipative context in Floquet form

Let us consider the dissipative context with the usual local coordinates[15] and integrable family of vector fields.

$$X^\mu = f(z,\mu)\frac{\partial}{\partial x} + g(z,\mu)\frac{\partial}{\partial z}, \tag{3.35}$$

where $g(0,0) = 0$ and $f(0,\mu) = \omega(\mu)$. Generally $D_z g(0,0)$ is a non-singular matrix, which implies that $g(z,\mu) = 0$ can be by solved the implicit function theorem with respect to z as $z = a(\mu)$ with $a(0) = 0$. Thus, there exists a function $\mu \mapsto a(\mu)$ such that $g(a(\mu),\mu) \equiv 0$. This allows us to construct the coordinate transformation $y = z - a(\mu)$.

After applying this coordinate transformation to the vector fields (3.35), and subsequently substituting the Taylor expansion of f and g in the y variable about the point 0, we obtain

$$X^\mu = \left(\omega(\mu) + O(y)\right)\frac{\partial}{\partial x} + \left(\Omega(\mu)y + O(y^2)\right)\frac{\partial}{\partial y}, \tag{3.36}$$

which is the *Floquet* form of an integrable family of vector fields in dissipative context. And, where we have that $\Omega(\mu) = D_y g(0,\mu)$ is non-singular for all μ.

Non-Floquet KAM theory

KAM theory does exist for certain non-Floquet tori, i.e. where the matrix Ω depends on x, such as hyperbolic lower-dimensional tori in Hamiltonian context. This involves constructing an additional diffeomorphism that reduces such tori to a Floquet form, which requires additional conditions. This is unfortunately beyond the scope of this thesis, therefore moving forward from here on only integrable families of vector fields that are reducible to the Floquet form are considered.

[14] Two complex eigenvalues $a \pm ib$ are conjugated, a normal frequency is defined as only the positive value b with the imaginary i dropped.

[15] The variable y relabelled as z.

3.5 Parameters and frequency control

This section discusses the use of parameters and how they control the frequencies of the invariant tori.

The frequency map is defined, which governs the use of parameters to control the frequencies, and is subject to a non-degeneracy condition that is required for KAM theory to yield meaningful results. The section also discusses the internal and external parameters, and briefly mentions the reduction of parameters, which is a major topic in KAM literature. Lastly, Kolmogorov's and the iso-energetic non-degeneracy conditions are presented and discussed with respect to the abstract non-degeneracy condition.

Frequency map

The persistence of each quasi-periodic invariant torus mainly depends on the frequencies of the torus. If the frequencies satisfy the Diophantine conditions[16], then the torus persists under small perturbations. The frequencies are, in turn, determined by the vector fields in the integrable family $\{X^\mu\}_\mu$, and care should be taken to ensure that a sufficient number of these vector fields do, in fact, result in a Diophantine torus.

To this end, it is useful to study the so-called frequency map of the tori, which is given by the x-component of the vector fields $\{X^\mu\}_\mu$. In local coordinates, this is defined as follows.

Definition 3.5.1 (Frequency map)**.** Let X^μ be as in (3.36), then the frequency map is given by

$$\omega : \mathbb{R}^s \to \mathbb{R}^n : \mu \mapsto \omega(\mu). \tag{3.37}$$

In lower-dimensional cases, where normal frequencies are present, these must be controlled too. Normal-internal near-resonances must be excluded as these also result in small denominators. Normal frequencies are therefore subject to arithmetic condition analogous to the internal frequencies, and analogous non-degeneracy conditions are required.

Internal and external parameters

It is possible to define a secondary parameter map that determines the parameters μ.

$$\mu : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^s : (y, \lambda) \mapsto \mu(y, \lambda). \tag{3.38}$$

The parameters λ can be freely used and are referred to as the *external* parameters. The number of external parameters available is theoretically unlimited. However, in practice, this number is usually restricted depending on the context and author[17].

In some contexts, particularly in the Hamiltonian context, the variables y can also be used as variables in the parameter map μ, and in extension the frequency map ω. Variables used in this way are referred to as the *internal* parameters. For the variables y to be used as internal parameters, the resulting parameters $\mu(y, \lambda)$ must still define an integrable family of vector fields. In other words, at each value of y in a small neighborhood of the origin, the vector field $X^{\mu(y,\lambda)}(\cdot, y)$ must be integrable.

This is quite restrictive, and not all variables can be internal parameters. The number of internal parameters available is heavily dependent on the context of the KAM theory.

Examples

For example, in the general dissipative context the only restriction on the dimension of y is non-negativity, and it could be zero. Directly implying that there are no internal parameters. Moreover, and more importantly, in general integrable dissipative systems the invariant tori are usually isolated. This implies that in the neighbourhood of T^n no other invariant tori exist. Thus varying the variables y over such a neighbourhood does not result in an integrable family of vector fields, which in turn means that the variables y cannot be used as internal parameters.

On the other side of the spectrum, we have the Hamiltonian context, where under the flow of integrable vector fields it is very common to have "shells" of invariant tori surrounding the invariant torus at $y = 0$. These are naturally parameterized by the variables y, thus allowing us to use the variables y as an internal parameters.

[16] Or any other suitable arithmetic conditions.

[17] KAM theorems with the least number of external parameters are more widely applicable and thus more desirable.

3.5.1 Standard non-degeneracy condition

As stated earlier, parameterized KAM theory aims to determine for which parameter values of μ the torus survives under a perturbation[18].

In the previous subsection, we saw that the parameters determine the frequency, and in section 2.2, it was discussed that the frequencies ω determine if an invariant torus survives. If the frequencies satisfy the Diophantine conditions, the corresponding invariant torus survives small perturbations. Therefore, the obvious method of determining for which parameters the invariant tori survive is to pull the set of frequencies that satisfy the Diophantine conditions back into the parameter space by means of the frequency map.

For this pull-back to fully transfer the structure of the Diophantine frequency set and, more importantly, its properties regarding its measure, the frequency map is required to be submersive. This is the standard[19] non-degeneracy condition in KAM theory.

Definition 3.5.2 (Standard non-degeneracy condition)**.** Let $\{X^\mu\}_\mu$ be an integrable family of vector fields. If the frequency map $\omega(\mu)$ is submersive near $\mu = \mu_0$ then $\{X^\mu\}_\mu$ is said to be *non-degenerate* at μ_0.

For a formal proof that this is indeed a sufficient condition, consult [51, 62, 72, 86].

The standard non-degeneracy condition ensures that the structure and properties of the standard Diophantine frequencies can be successfully pulled back into the parameter domain. This, in turn, implies that the parameters for which the tori survive have a Cantorized structure in the full parameter space and have a positive measure depending on the chosen constants γ and τ. Thus, it can be concluded that *many* invariant quasi-periodic tori survive.

Parameter reduction

Not all available internal and external parameters have to be used, and it is often desirable to minimize the number of parameters as much as possible. Most KAM theorems aim to use the least number of external parameters possible, as low-parameter[20] theorems are more flexible in their application.

For some contexts and integrable families of vector fields, external parameters are necessary to prove meaningful persistence. Determining the minimum number of parameters required for a given context and subsequently studying the degenerate cases, where the available parameters are below the current[21] minimum, is an active field of KAM theory.

Another notable motivation for the reduction of parameters is the study of lower-dimensional invariant tori in the Hamiltonian context. The internal frequencies can be fully controlled by internal parameters because, for an n-dimensional torus with n internal frequencies, there are n action variables. However, each additional codimension of the torus results in an additional normal frequency that must also be controlled, without any additional internal parameters.

The aim was to reduce the amount of parameters required to control the internal frequencies, so that the excess internal parameter could be used to control the normal frequencies. Thus reducing the amount of required external parameters to prove a KAM theorem for lower-dimensional tori.

The effort to decrease the required parameters in the Hamiltonian context has since been perfected by the Rüssmann non-degeneracy condition [79, 81], which allowed for KAM theory on the persistence of lower-dimensional tori in Hamiltonian systems without the need for external parameters. This has since been applied to prove the persistence of entire quasi-periodic bifurcation scenarios in Hamiltonian systems without the need for external parameters, for examples consult the book [38] by Hanßmann.

[18] In this section, we only use the generic parameters μ. Internal and external parameters can be mapped to this parameter by means of a secondary map, as given in (3.38).

[19] This non-degeneracy condition is also commonly referred to as the BHT non-degeneracy condition, named after Broer, Huitema, and Takens, who first stated it in this form.

[20] Also referred to as miniparameter KAM theorems.

[21] With each new non-degeneracy condition that requires fewer parameters, the minimum number of parameters decreases.

3.5.2 Non-degeneracy conditions for Hamiltonian context

In this subsection, we restrict ourselves to the Hamiltonian context. As mentioned before, KAM theory in the Hamiltonian context is remarkably nice since it allows for n internal parameters. Furthermore, since the frequency is also n-dimensional, it is quite possible to construct a non-degenerate (internal) frequency map $y \mapsto \omega(y)$ without the need for external parameters. As such, the Hamiltonian context has the most extensive and interesting non-degeneracy conditions. However, in this section, only three examples are given.

For more non-degeneracy conditions, including those for lower-dimensional tori and normal frequencies, please refer to the study by Hanßmann [39] or the survey by Rüssmann [81].

We consider the integrable family of Hamiltonian vector fields $\{X^y\}_y$ parameterized by the internal parameter y, with each X^y defined as follows

$$X^y = \omega(y)\frac{\partial}{\partial x}. \tag{3.39}$$

Kolmogorov's non-degeneracy condition

This is the original non-degeneracy condition for KAM theory in the Hamiltonian context, as stated by Kolmogorov in his 1954 study [53]. It is quite straight-forward on what it achieves.

Definition 3.5.3 (Kolmogorov non-degeneracy condition)**.** Let $\{X^y\}_y$ be as above. If the frequency map $y \mapsto \omega(y)$ is a local diffeomorphism at $y = 0$, or equivalently $\det D\omega(y) \neq 0$ for $y = 0$, then $\{X^y\}_y$ satisfies the *Kolmogorov non-degeneracy condition.*

Since ω is a local diffeomorphism at $y = 0$, it is also submersive at $y = 0$, thus satisfying the standard non-degeneracy condition (see definition 3.5.2).

Iso-energetic non-degeneracy condition

This condition is due to Arnol'd [1] and is referred to as the Arnol'd non-degeneracy or "relaxed" non-degeneracy condition. It requires one less parameter than the Kolmogorov condition, and has been used to state "relaxed" KAM theorems in [17, 18].

This is useful in applications where the parameter is needed for other purposes, such as an unfolding parameter in simple bifurcation scenarios.

Definition 3.5.4 (Iso-energetic non-degeneracy condition)**.** Let $\{X^y\}_y$ be as above. If the frequency map $(c, y) \mapsto c\omega(y)$ is submersive at $y = 0$, then $\{X^y\}_y$ satisfies the *iso-energetic non-degeneracy condition.*

For a Diophantine frequency ω and some constant $c > 1$, we have that $c\omega$ is also a Diophantine frequency. This allows us to map y to an S^{n-1} in $\mathbb{R}^n$ and, by multiplying by c, still reach all Diophantine frequencies. The multiplication by c is usually achieved by adjusting the time scale of the system.

Rüssmann's non-degeneracy condition

The last condition that is mentioned in this thesis is the Rüssmann non-degeneracy condition[22]. This condition can be stated for an arbitrary high order $L \in \mathbb{N}$, thus allowing for an arbitrary reduction of parameters required for a sufficient non-degeneracy condition.

Definition 3.5.5 (Rüssmann non-degeneracy condition)**.** Let $\{X^y\}_y$ be as above. If the frequency map $y \mapsto \omega(y)$ satisfies

$$\operatorname{span}\left\{\frac{\partial^{|l|}\omega}{\partial y^l} : |l| \leq L\right\} = \mathbb{R}^n, \tag{3.40}$$

for $L \in \mathbb{N}$, then $\{X^y\}_y$ satisfies the *Rüssmann non-degeneracy condition.*

This condition follows from Bakhtin's theorem (see theorem 2.2.7), which states that there is a sufficient number of Diophantine frequencies in a submanifold of the frequency space $\mathbb{R}^n$ if the submanifold is sufficiently "bent". The Rüssmann non-degeneracy condition exactly implies this by requiring that the frequency map ω, which defines a submanifold in $\mathbb{R}^n$, satisfies (3.40). For more detailed discussion, consult [79, 81] and [17].

Lastly, we note that for $L = 1$ Rüssmann's non-degeneracy condition is equivalent to the iso-energetic non-degeneracy condition.

[22] Please note that this non-degeneracy condition should not be confused with the Bruno-Rüssmann arithmetic conditions.

3.6 Complete Hamiltonian Lagrangian context

This thesis only fully considers KAM theory for a specific case of the Hamiltonian isotropic context, therefore only this context is fully examined. Many of the more difficult contexts have additional properties that would have to be taken into account and the associated KAM theorems are therefore more complex, and unfortunately outside the scope of this thesis.

Local coordinates

In this context, we assume that there exists an integrable s-dimensional family of Hamiltonian vector fields with n degrees of freedom on a $2n$-dimensional exact symplectic manifold, with the exact symplectic form ω^2, and where the invariant torus T^n has n dimensions. Lastly, we assume T^n is isotropic and, therefore, in this case Lagrangian.

For any invariant torus, it is possible to define a coordinate system $(\theta \in T^n, w \in \mathbb{R}^n)$ on M such that T^n is defined by $\{w = 0\}$. However, these coordinates are not ideal for KAM theory, and it is convenient to apply change of coordinates of the type below.

$$\theta = x + Ay, w = By \quad \text{with} \quad x \in T^n,\ y \in \mathbb{R}^n, \tag{3.41}$$

where A, B are constant matrices. After applying a cleverly chosen transformation of this type, the manifold has the familiar structure

$$M = T^n \times \mathbb{R}^n, \tag{3.42}$$

and the symplectic form has the canonical form

$$\omega^2 = \sum_{i=1}^{n} dx_i \wedge dy_i. \tag{3.43}$$

These are the new local coordinates in which KAM theory in the Lagrangian context is stated. It is often assumed, without loss of generality, that the manifold and symplectic form already have this form.

The integrable family of vector fields

For integrable s-parameter families of Hamiltonian vector fields $\{X^\mu\}_\mu$, on M with ω^2 as above, each X^μ has the form

$$X^\mu = \frac{\partial N^\mu}{\partial y}\frac{\partial}{\partial x} - \frac{\partial N^\mu}{\partial x}\frac{\partial}{\partial y} = \frac{\partial N^\mu}{\partial y}\frac{\partial}{\partial x}, \tag{3.44}$$

where $N^\mu(y)$ the Hamiltonian functions associated with the vector fields X^μ.

The vector fields, and in extension the Hamiltonian functions N^μ, are x-independent by the definition of integrability. Therefore, we have $\frac{\partial}{\partial x}N^\mu = 0$ and x can be omitted from N^μ. The Hamiltonian functions N^μ have the following form

$$N^\mu(y) = h(y, \mu) = \langle \omega(\mu), y \rangle + O(y^2), \tag{3.45}$$

and their associated vector fields X^μ are in the form

$$X^\mu = (\omega(\mu) + O(y))\frac{\partial}{\partial x}, \tag{3.46}$$

where the y-component of the vector field is zero, and thus omitted.

Simplified form

In KAM theory, where y is usually restricted to a very small neighbourhood of the origin, the higher order terms of h are small enough to be interpreted as a perturbation and thus usually added to the perturbation term. This allows us to use an even more well-behaved and understood system as the integrable system in KAM theorem statements. In practical terms, the simplified systems lie close enough to the original integrable systems that their perturbations largely overlap and thus almost the same conclusions can be made.

As this simplifies the calculations required to prove the theorem, it is very common to restrict the form of the integrable Hamiltonians N^μ and integrable vector fields X^μ to

$$N^\mu(y) = \langle \omega(\mu), y \rangle \quad \text{and} \quad X^\mu = \omega(\mu)\frac{\partial}{\partial x}. \tag{3.47}$$

Vector fields in this form can also be obtained by means of normal linearization, which is a more robust method to obtain such simplified vector fields. Unfortunately, this is outside the scope of this thesis. For further information, see [18] or [17].

3.7 KAM method overview

At this point, we have discussed the context of KAM theory, integrable vector fields, and the Lie algebra method to construct structure-preserving diffeomorphisms. This last section is written to demonstrate the construction of the conjugation and explain the iteration scheme used in KAM theory.

Let X be an integrable vector field on M, and let $\tilde{X}$ be a small perturbation of X. KAM theory aims to construct a conjugation

$$\Phi : M \to M \quad \text{such that} \quad \Phi_* \tilde{X} = X^{\omega_\infty}, \tag{3.48}$$

where X^{ω_∞} is an integrable vector field, which may not be equal to X. This conjugation must be smooth, preserve the structure defined on the manifold, and does not necessarily exist for all integrable vector fields and perturbations.

Goal of KAM theory

KAM theory constructs the conjugation Φ as the limit of a sequence of cleverly chosen smooth structure-preserving transformations. Each successive transformation transforms the perturbed system to a "less" perturbed system. The less perturbed system is a perturbation of an integrable system that has a smaller non-integrable part than the non-integrable part of the previous perturbed system. As the number of these transformations applied to a perturbed system approaches infinity, the resulting less perturbed system approaches a fully integrable system.

Constructing appropriate structure-preserving transformations that minimize the non-integrable part and formally proving this is one of the main difficulties in KAM theory.

Fortunately, there is quite some flexibility in which transformations sufficiently shrink the non-integrable part. Therefore, they can be constructed with additional considerations, such as adherence of the new integrable part to specific forms that go beyond just integrability[23]. For example, in finite differentiable KAM theory, each of these transformations has been adapted with some degree of smoothing to control the loss of differentiability with each step.

In this section, we discuss the general method of constructing these transformations. We do not impose any requirements on the transformation beyond that they result in smaller non-integrable parts.

Context of this demonstration

For this example, we assume the classical Hamiltonian isotropic $(n, 0, 0)$ context, otherwise known as the Lagrangian context.

Let the manifold be given by $M = \mathbb{T}^n \times \mathbb{R}^n$ and let M be equipped with the local coordinates and the canonical symplectic form, as described in section 3.6. Let $\mathbb{Y} \subset \mathbb{R}^n$ be a small neighbourhood of the origin, and define the domain $D = T^n \times \mathbb{Y}$. Let $\mu \in \mathbb{R}^n$ and let $\{X^\mu\}_\mu$ be an n-parameter integrable family of Hamiltonian vector fields on D given by

$$\{X^\mu\}_\mu \quad \text{such that} \quad X^\mu = \omega(\mu)\frac{\partial}{\partial x}. \tag{3.49}$$

Here, we use the simplified vector field as described in section 3.6, and as before we omit the zero y-component of X. To simplify the following discussion and calculations further, we decouple the frequencies from the parameter μ by introducing an externally given frequency $\omega \in \mathbb{R}^n$. The frequency map and associated non-degeneracy condition are rendered trivial by this action. This results in a rather simple family given by

$$\{X^\omega\}_\omega \quad \text{such that} \quad X^\omega = \omega\frac{\partial}{\partial x}. \tag{3.50}$$

Associated with these vector fields are the Hamiltonian functions $N(y, \omega) = \langle\, \omega,\, y \,\rangle$.

KAM theory considers small perturbations P of the Hamiltonian functions N, resulting in small perturbations of the vector field X^ω. The Hamiltonian of these vector fields is given by

$$H(x, y, \omega) = N(y, \omega) + P(x, y, \omega). \tag{3.51}$$

In the next few sections, we continue with a more general method using the perturbed vector fields instead of the Hamiltonian functions. This method can be more readily adapted to the other contexts. The perturbed vector fields themselves are given by

$$\tilde{X}(x, y, \omega) = (\omega + f(x, y, \omega))\,\frac{\partial}{\partial x} + g(x, y, \omega)\frac{\partial}{\partial y}, \tag{3.52}$$

where f and g are small perturbation given by $\partial_y P$ and $-\partial_x P$ respectively.

[23] For example, a specific normal form.

3.7.1 KAM iterative scheme

As was said before, the equation (3.48) is very difficult to solve directly, and when applying the KAM method we manage these by iteratively approximating the conjugation by coordinate transformations, where each coordinate transformation transforms the perturbed system closer to an integrable one.

We start by defining Φ in a more formal way.

$$\Phi : T^n \times R^n \times R^n \to T^n \times R^n \times R^n \quad \text{such that} \quad \Phi_* \tilde{X}^{\tilde{\omega}} = X^{\omega_\infty}, \tag{3.53}$$

Note that the frequencies $\tilde{\omega}$ and ω_∞ are not necessarily the same. We define the sequence of transformations $(\Phi_j)_j$ by

$$\Phi_{j+1} : D_{j+1} \to D : \Phi_{j+1} := \Phi_j \circ \Psi_j \quad \text{where} \quad \Phi_0 = \text{id}, \tag{3.54}$$

where the transformations Ψ_j are defined by

$$\Psi_j : D_{j+1} \to D_j \quad \text{such that} \quad (\Psi_j)_* \tilde{X}_j^{\omega_j} = \tilde{X}_{j+1}^{\omega_{j+1}}. \tag{3.55}$$

Thus, we have $(\Phi_j)_* \tilde{X}^{\tilde{\omega}} = \tilde{X}_j^{\tilde{\omega}_j}$. Our goal is to construct each Ψ_j in a way that satisfies three criteria: firstly, Ψ_j must be well-defined; secondly, $\Phi_\infty = \lim_{j \to \infty} \Phi_j$ must converge; and lastly, Φ_∞ must satisfy the conjugation relation (3.53). The form of each $\tilde{X}_j^{\omega_j}$ is given by

$$\tilde{X}_j^{\omega_j} = (\omega_j + f_j(x_j, y_j, \omega_j)) \frac{\partial}{\partial x} + g_j(x_j, y_j, \omega_j) \frac{\partial}{\partial y} = X^{\omega_j} + R_j, \tag{3.56}$$

where X^{ω_j} and R_j are the integrable and non-integrable parts, respectively. If we can find a convergent sequence Φ_j such that the non-integrable part R_j approaches zero as $j \to \infty$, then we have found the desired conjugation between the integrable and perturbed vector field.

Iteration schemes

In the construction of each Φ_j, small denominators occur and must be managed. There are two ways to handle these small denominators when applying the traditional KAM method.

- *Kolmogorov's scheme*: This is the original solution to the problem by Kolmogorov [53]. The scheme fixes ω to a single Diophantine frequency and allows us to use all Diophantine conditions at each step of the iteration. Due to the fixing of ω to a single frequency, Kolmogorov's scheme only determines the persistence of a single KAM torus.
- *Arnol'd's scheme*: This scheme was developed by Arnold in his original study [1]. This scheme confines the frequencies ω to shrinking domains at each KAM step. On these shrinking domains, the frequencies ω satisfy the Diophantine conditions for $|k|$ below an increasingly higher cut-off d. This is sufficient for the construction of Ψ. This scheme was also demonstrated in the one-bit problem solution in section 2.4.

In this demonstration, we continue with Arnold's scheme.

Plus (+)-notation

The (+)-notation convention is commonly used in many KAM proofs, including this thesis. When dealing with a single iteration step, indices $j+1$ and j are often dropped in favor of the symbol $+$ and nothing. For instance, we write ω instead of ω_j, ω_+ instead of ω_{j+1}, and so on. Similarly, the notation includes the vector fields $\tilde{X}_j^{\omega_j}$ and their components f_j.

However, there is one notable exception to this convention in this thesis, which concerns the x and y variables. Since, we use only x_{j+1} and y_{j+1} in the construction and we write them as x and y.

Construction of Ψ

In the next sections, we discuss the construction of a suitable map

$$\Psi : D_+ \to D : (x, y, \omega_+) \mapsto (\psi_\omega(x, y), \omega) \quad \text{such that} \quad \Psi_* \tilde{X}^{\omega} = \tilde{X}_+^{\omega_+}, \tag{3.57}$$

which links Φ to Φ_+ by $\Phi_+ = \Phi \circ \Psi$. Since this demonstration is set in the Hamiltonian context, we require Ψ and, by extension, ψ_ω to be symplectic. The shrinking domains D_+ and D are not the focus of this demonstration and are therefore not further specified.

The construction of Ψ consists of two parts:

Firstly, the construction of the symplectic map ψ_ω. This is the more challenging part since it involves many additional steps, both for convenience and necessity.

Secondly, the construction of the parameter shift $\omega_+ \mapsto \omega$. This is a relatively straightforward application of the implicit function theorem.

3.7.2 Construction of ψ_ω

We define ψ_ω as the time-one flow map of a Hamiltonian vector field V, i.e. $\psi_\omega = \varphi_V^1$. This is by far the most convenient way to construct ψ_ω as it satisfies one of the requirements by being a symplectic map by lemma 3.2.10. Moreover, this approach allows us to use the Lie algebra-based method described in section 3.3 to calculate the vector field V.

The construction of the vector field V consists of four parts:

Firstly, we must "linearize" the conjugation relation. Directly solving the conjugation relation directly is too complicated, and several simplifications can be made to still obtain a sufficiently simplified conjugation relation. Secondly, we must linearize and truncate the non-integrable part. Solving the simplified conjugation relation still involves small denominators, and these require additional steps to manage. Thirdly, by choosing an appropriate new integrable part, we can solve the vastly simplified conjugation relation to obtain the desired ψ_ω. Lastly, we derive the equation for the new non-integrable part.

Afterwards, we briefly discuss all important equations in the Hamiltonian function notation.

This section only covers the equations that need to be solved to obtain ψ_ω, so no explicit calculations are made. The shrinking domains required to manage the small denominators and the estimates on the new non-integrable part R_+ are not considered in this section to maintain a transparent construction.

Approximate conjugation relation

The conjugation relation $(\psi_\omega)_* \tilde{X}^\omega = \tilde{X}_+^{\omega_+}$ in Lie algebra terms (see section 3.3.5) is given by

$$(\psi_\omega)_* \tilde{X}^\omega = \mathrm{Ad}_{\psi_\omega}(\tilde{X}^\omega) = \tilde{X}_+^{\omega_+}, \tag{3.58}$$

which is equivalent by (3.26) to

$$\tilde{X}_+^{\omega_+} = \tilde{X}^\omega + [V, \tilde{X}^\omega] + \int_0^1 (1-t) \cdot [V, [V, \tilde{X}^\omega]]_{\varphi_V^t} \, dt. \tag{3.59}$$

This equation is usually too difficult[24] to solve directly for V, and as such, we must approximate the solution.

Firstly, we assume that the norm of V is very small because ψ_ω must be close to the identity, which follows from the very small difference between X and $\tilde{X}$. This allows us to drop the integral, as this term is of the order of V^2. This step is the first linearization of the conjugation relation. The negligibility of the integral is explicitly verified later on by estimating the new non-integrable part, of which the integral term is a part.

We drop the integral and substitute $\tilde{X}_+^{\omega_+} = X^{\omega_+} + R_+$ and $\tilde{X}^\omega = X^\omega + R$ to obtain

$$X^{\omega_+} + R_+ = X^\omega + R + [V, X^\omega + R] = X^\omega + R + [V, X^\omega] + [V, R]. \tag{3.60}$$

This equation can often be further simplified. Since $\tilde{X}^\omega$ is presumed to be a small perturbation of the integrable system X^ω, the non-integrable part R is assumed to be of the same order of smallness as V. Therefore, the term $[V, R]$ should be of the order of V^2 and can be safely dropped. In addition, we assume that the new non-integrable part R_+ is very small and can be dropped as well. Again, the negligibility of these terms is explicitly verified later by estimating the new non-integrable part.

This results in the final "linearized" conjugation equation

$$X^{\omega_+} = X^\omega + R + [V, X^\omega]. \tag{3.61}$$

The omissions of $[V, R]$ and R_+ are technically not a linearizations, but often this distinction is not made. Therefore, within the Hamiltonian context, the equation (3.61) is usually also referred to as the linearized conjugation relation.

The new integrable part X^{ω_+} in (3.61) is not uniquely determined and we have some[25] freedom in choosing one. Later in this section, it is set to a specific value, which is simple for the calculations but still sufficient for estimating the new non-integrable part.

All errors generated by this approximation of the conjugation relation and the potentially non-optimal choice of the new integrable part X^{ω_+} are collected in the new non-integrable part R_+. These are eventually proven to be sufficiently small for the KAM method to work.

[24] Very well-behaved perturbations could have direct solutions and thus result in an integrable $\tilde{X}+^{\omega+}$ in a single step.

[25] The new integrable part must still result in a solvable linearized conjugation relation, and this equation must result in a sufficiently small R_+.

Linearization of R

The equation (3.61) is still not readily solvable. When attempting to naively solve (3.61), one encounters small denominators, which must be managed.

Before the truncation, another simplification to (3.61) is made. Recall that the domain of ψ_ω in the y variable is very close to the origin. This allows us to linearize the non-integrable part R about the origin in the y variable and only consider its lower order terms to further simplify the equation. The smaller higher order terms are added to the new non-integrable part.

The linearization is formalized by the following operator.

Definition 3.7.1 (Linearization operator)**.** Let $F(x,y,\omega) = f(x,y,\omega)\frac{\partial}{\partial x} + g(x,y,\omega)\frac{\partial}{\partial y}$, we define the linearization operator as follows

$$F_{\text{lin}}(x,\omega) := f(x,0,\omega)\frac{\partial}{\partial x} + g(x,0,\omega)\frac{\partial}{\partial y}. \tag{3.62}$$

With this operator, we can simplify (3.61) to

$$X^{\omega_+} = X^\omega + [V, X^\omega] + R_{\text{lin}}. \tag{3.63}$$

Truncation of R

Solving (3.63) still involves small denominators, which the two schemes deal quite differently with.

In Kolmogorov's scheme, the frequency ω remains constant over all iterations. Therefore, if this frequency is Diophantine, the equation (3.63) can be solved without the need for Fourier truncation.

In Arnol'd's scheme, at each iteration, the frequency ω is part of a shrinking domain, on which each ω has the property

$$|\langle k, \omega\rangle| \geq \tfrac{1}{2\gamma}|k|^\tau \quad \text{for each} \quad 0 < |k| < d \quad (k \in \mathbb{Z}^n). \tag{3.64}$$

Thus, the small denominators are sufficiently large on the domain for k below some ultraviolet cut-off $d \in \mathbb{N}$. The shrinking domains are defined in such a way that d increases at each iteration, thus making this estimate (3.64) closer and closer to the Diophantine conditions. We continue using Arnol'd's scheme as it is more commonly used in modern KAM theory literature.

As we can only control small denominators up to d, we can only solve for Fourier series truncated at the cut-off. Therefore, we truncate the Fourier series of R at d. This Fourier truncation is formalized by the following operator.

Definition 3.7.2 (Fourier truncation operator)**.** Let $d \in \mathbb{N}$ and let $f(x,\omega)$ be some function on $T^n \times \mathbb{R}^n$ and consider its Fourier series $f(x,\omega) = \sum_{k\in\mathbb{Z}^n} f^k(\omega)e^{i\langle k,x\rangle}$. We define the Fourier truncation operator as follows

$$f_d(x,\omega) = \sum_{|k|\leq d} f^k(\omega)e^{i\langle k,x\rangle}. \tag{3.65}$$

With this operator, we can simplify (3.63) to

$$X^{\omega_+} = X^\omega + [V, X^\omega] + R_{\text{lin},d}, \tag{3.66}$$

which is usually referred to as the truncated linearized conjugation relation[26]. In equation (3.66), we finally have a solvable equation for the vector field V and, in extension, the transformation ψ_ω.

Solving for vector field V

The formal solution of the equation (3.66) is omitted. A slight note on the form of V is still given. The vector field has the form

$$V = u(x,\omega)\frac{\partial}{\partial x} + v(x,\omega)\frac{\partial}{\partial y}, \tag{3.67}$$

where functions u and v have a Fourier series that is cut-off at d. The form of V is determined by the form of $R_{\text{lin},d}$, which in turn follows from the form of the perturbation (cf. (3.52)). The vector field V inherits this form via the Taylor and Fourier series-based solution of equation (3.66).

[26] Or, as mentioned before, the truncated homological equation.

New non-integrable part

A lot of approximations and simplifications were made to obtain equation (3.66), and these inaccuracies all contribute to the new perturbation term R_+. Here we give the equation that determines R_+ for our chosen X^{ω_+} and transformation ψ_ω constructed with the vector field V which satisfies (3.66).

Suppose we have solved equation (3.66) for V and obtained $\psi_\omega := \varphi_V^1$. Then we have

$$\mathrm{Ad}_{\psi_\omega}(X^\omega + R) = \mathrm{Ad}_{\psi_\omega}(X^\omega) + \mathrm{Ad}_{\psi_\omega}(R_{\mathrm{lin},d}) + \mathrm{Ad}_{\psi_\omega}(R - R_{\mathrm{lin},d}) \tag{3.68}$$

$$= X^\omega + [V, X^\omega] + \int_0^1 (1-t)[V,[V,X^\omega]] \circ \varphi_V^t \, dt \tag{3.69}$$

$$+ R_{\mathrm{lin},d} + \int_0^1 [V, R_{\mathrm{lin},d}] \circ \varphi_V^t \, dt + \mathrm{Ad}_{\psi_\omega}(R - R_{\mathrm{lin},d}). \tag{3.70}$$

We set $\mathrm{Ad}_{\psi_\omega}(X^\omega + R) = X^{\omega_+} + R_+$ with our chosen new integrable part X^{ω_+}, and subtract (3.66) to obtain

$$R_+ = \int_0^1 [V, (1-t)[V,X^\omega] + R_{\mathrm{lin},d}] \circ \varphi_V^t \, dt + \mathrm{Ad}_\psi(R - R_{\mathrm{lin},d}). \tag{3.71}$$

The new non-integrable part R_+ can be proven to be smaller than the previous R, implying that each linking of Ψ brings Φ_+ closer to the desired conjugation Φ_∞.

3.7.3 Parameter shift $\omega^+ \mapsto \omega$

Lastly, in Arnol'd's scheme, the new integrable part is chosen to accommodate the solution to the linearized conjugation relation.

New integrable part

When solving the linearized conjugation relation, the zero-th order terms of R must be absorbed by the new integrable part X^{ω_+}. The vector field V cannot absorb these values due to the derivative in the definition of the Lie bracket.

Therefore, we choose the new integrable part X^{ω_+} to be X^ω with these leftover zero-th order Fourier coefficients of R absorbed. The simplest way to add these is to simply add the average of R over x to X^{ω_+}. Other possibilities exist but these only serve to complicate the matter. To precisely define X^{ω_+}, we introduce the average operator, which is defined as follows.

Definition 3.7.3 (Average operator)**.**

$$[R(\cdot,0,\omega)] = [f(\cdot,0,\omega)]\frac{\partial}{\partial x}, \tag{3.72}$$

where $[f(\cdot,0,\omega)]$ is given by

$$[f(\cdot,0,\omega)] = \frac{1}{\mathrm{Vol}(T^n)} \int_{T^n} f(x,0,\omega)\, dx. \tag{3.73}$$

With this operator we set the new integrable part to

$$X^{\omega_+} = X^{\omega+\Lambda(\omega)} = X^\omega + [R(\cdot,0,\omega)] = (\omega + [f(\cdot,0,\omega)])\,\frac{\partial}{\partial x} = (\omega + \Lambda(\omega))\,\frac{\partial}{\partial x}, \tag{3.74}$$

where $\Lambda(\omega) := [f(\cdot,0,\omega)]$ is referred to as the parameter shift. We see that with each iteration we shift the parameter ω, except in the case that $[f(\cdot,0,\omega)] \equiv_\omega 0$.

Parameter shift

We have that our chosen new integrable part $X^{\omega+\Lambda(\omega)}$, as defined in (3.74), is not in the normal form. However, as this is required for the construction of each transformation ψ_ω, the new integrable part must be transformed back into the normal form before continuing with the construction of the next transformation Ψ_+.

We see that, by equation (3.74), that we do have a map $\omega \mapsto \omega_+$ given by

$$\omega_+ = \omega + [f(\cdot, 0, \omega)] = \omega + \Lambda(\omega). \tag{3.75}$$

Since $\Lambda(\omega)$ is small due to it being derived from the small perturbation f, the map (3.75) is close to the identity. Hence, we can apply the inverse function theorem to show that there exists an inverse map $\phi : \omega_+ \mapsto \omega + \Lambda(\omega)$. Applying the map ϕ to

$$(\psi_\omega)_*(\tilde{X}^\omega) = (\psi_\omega)_*(X^\omega + R) = X^{\omega+\Lambda(\omega)} + R_+, \tag{3.76}$$

results in the actual new integrable part and new perturbation term, given by

$$\phi_*(X^{\omega+\Lambda(\omega)} + R_+) = X^{\omega_+} + (R_+ \circ \phi) = \tilde{X}_+^{\omega_+}, \tag{3.77}$$

where X^{ω_+} is in normal form.

Thus, the transformation defined by $\Psi : (x, y, \omega_+) \mapsto (\psi_\omega(x, y), \phi(\omega_+))$ satisfies $\Psi_* \tilde{X}^\omega = \tilde{X}_+^{\omega_+}$.

3.7.4 Construction of Ψ: Hamiltonian functions

In the Hamiltonian context, it is usually far easier to make all calculations and estimates for the vector fields using their corresponding Hamiltonian functions. Let $H = N + P$ be an analytic perturbation of an integrable Hamiltonian N, and let $H_+ = N_+ + P_+$ be the less perturbed system.

The conjugation by a transformation ψ_ω in Hamiltonian functions is given by

$$H \circ \psi_\omega = H_+ = N_+ + P_+. \tag{3.78}$$

When using the Lie algebra method ψ_ω is defined as the time-one flow map of the vector field induced by the Hamiltonian F. The full conjugation relation in Hamiltonian function is given by

$$H \circ \psi_\omega = H + \{H, F\} + \int_0^1 (1-t)\{\{H, F\}, F\}\, dt. \tag{3.79}$$

This is directly analogous to the conjugation relation (3.59) for vector fields, and analogous linearizations and simplifications can be applied to obtain

$$N_+ = N + \{N, F\} + P, \tag{3.80}$$

which is a direct analogue[27] to the relation (3.61). As in the previous sections, we can further simplify (3.80) and manage small denominators by linearizing and truncating the non-integrable part P. This results in the following equation

$$N_+ = N + \{N, F\} + P_{\text{lin},d}. \tag{3.81}$$

Solving the truncated linearized conjugation relation (3.81) for F with a suitably chosen N_+ gives us a sufficient approximate conjugation ψ_ω. Lastly, we state the equation for the new perturbation function in Hamiltonian functions

$$P_+ = \int_0^1 \{(1-t)\{N, V\} + P_{\text{lin},d}, V\} \circ \varphi_V^t\, dt + (P - P_{\text{lin},d}) \circ \psi_\omega, \tag{3.82}$$

which is again directly analogous to the vector field formulation (3.71).

Lastly, in the Hamiltonian function formulation, N_+ is very commonly chosen to be

$$N_+ = N + [P(\cdot, 0, \omega)], \tag{3.83}$$

where $[P(\cdot, 0, \omega)]$ is defined as

$$[P(\cdot, 0, \omega)] = \frac{1}{\text{Vol}(T^n)} \int_{T^n} P(x, 0, \omega)\, dx. \tag{3.84}$$

This is analogous to (3.74), and results in the parameter shift $\Delta(\omega) = [P(\cdot, 0, \omega)]$. From here on, the same methods are used to construct the parameter shift map ϕ as in the previous section.

[27] Within the Hamiltonian context, the equation (3.80) is also commonly referred to as the homological equation.

Part III: KAM proof

In this part, an example of a proof of the KAM theorem is presented. The theorem being proven is the classical KAM theorem in the classical Lagrangian context, stated below.

Theorem 4.0.1 (Classical KAM theory)**.** *Let $H = N + P$ be an analytic perturbation of a non-degenerate integrable Hamiltonian function N defined on $D := T^n \times \mathbb{Y}$. Then there exists a constant $\delta > 0$ such that if*

$$|P|_D = \sup_{(x,y)\in D} |P(x,y)| < \delta, \tag{4.1}$$

then all Diophantine quasi-periodic tori of the unperturbed system persist, being only slightly deformed. Moreover, they depend in a Whitney smooth way on ω and fill the phase space $T^n \times \mathbb{Y}$ up to a set of measure $O(\gamma)$.

We prove this classical KAM theorem by considering a specific case of a parameterized KAM theorem.

We decouple the frequencies ω from the action variables y, as this renders the non-degeneracy condition trivial and simplifies the calculation of the parameter shifts. We use the frequencies $\omega \in \Sigma \subset \mathbb{R}^n$, where Σ is a compact set, as the external parameters with the identity as the frequency map[1]. This corresponds to the vertical vector field X and the associated integrable Hamiltonian N given by

$$X^\omega = \omega \frac{\partial}{\partial x} \quad \text{with Hamiltonian function} \quad N(x,y,\omega) = \langle \omega, y \rangle. \tag{4.2}$$

Using this, we can state the decoupled or frequency parameterized KAM theorem.

Theorem 4.0.2 (Parameterized Hamiltonian KAM theorem.)**.** *Let $H = N + P$ be an analytic perturbation of an integrable Hamiltonian* (4.2)*, defined on a small neighbourhood D of $T^n \times \{0\} \times \Sigma$. Then there exist a constant $\delta > 0$ such that if*

$$|P|_D = \sup_{(x,y,\omega)\in D} |P(x,y,\omega)| < \delta \tag{4.3}$$

then there exists a Whitney regular diffeomorphism Φ on $T^n \times \mathbb{R}^n \times \Sigma$ such that

- *Φ is analytic and symplectic for fixed ω.*
- *Φ is C^∞-close to the identity.*
- *On $T^n \times \mathbb{R}^n \times \Sigma'_\gamma \cap \Phi^{-1}(T^n \times \{0\} \times \Sigma)$, the Hamiltonian function $H \circ \Phi = N_\infty$ is integrable.*

The proof of the above decoupled KAM theorem, given in this section, follows the traditional approach. The KAM method is applied with the standard Arnol'd iteration scheme.

After proving theorem 4.0.2, the classical KAM theorem can be easily derived from it. Hence, it is no surprise that decoupling the frequencies is a standard practice.

Outline of the part III

This part begins with a general outline and reading guide of the proof, as the KAM proof is extensive and can be challenging to follow without a good overview.

After the overview, we give the setting and statement of the parameterized KAM theorem. This includes the manifold, conditions, domain, norm, and the theorem statement. Defining the setting and necessary conditions is necessary for the formal statement of the KAM theorem. The setup, notation and proof itself is largely based upon KAM proofs found in quasi-periodic bifurcation theory, see [14, 15, 37, 38].

The proof of the parameterized KAM theorem is then divided into three sections. The first section is dedicated to the setup of the KAM proof. Here, we define the shrinking domains D_i and describe the construction of the sequence Φ_i using Arnol'd's scheme. The second section of a KAM proof is typically the largest and is dedicated to the KAM step. This section involves many estimates and is where the bulk of the proof lies. The final section of the KAM proof deals with the convergence of the conjugation and verifies all the desired properties stated in the KAM theorem.

Finally, we use theorem 4.0.2 to prove the classical KAM theorem 4.0.1.

[1] The identity map by default satisfies the standard non-degeneracy condition.

4.1 Outline of a KAM theorem proof

This is a standard outline of a KAM proof using the traditional KAM method. In this thesis, the proof of the KAM theorem is divided into three sections, besides the statement and prerequisites. While this division is not strictly necessary, it is helpful for the understanding of the proof.

The first step of a KAM theorem proof is the complete statement of the theorem.

- *The statement of the KAM theorem*: This involves specifying the situations in which the theorem can be applied, including the context, domain, normal forms of the integrable part, and so on.

This step can often be found as a standalone statement, as it adequately describes the KAM theorem for application purposes. The following three parts constitute the proof of the KAM theorem and are rarely found in isolation, as an incomplete proof is not very useful. They are:

- *The setup*: This is the first part of any KAM proof. Here, we define the iteration scheme and iterative domains, and corresponding sequences.
- *The KAM step lemma*: This lemma proves all the necessary properties for the construction of a single iteration step. This is usually the largest part of a KAM proof.
- *Convergence*: This last part proves the convergence of the scheme with the transformation constructed by the KAM step lemma. Lastly, we verify the properties of the resulting diffeomorphism.

The precise content of each of these parts is discussed below.

KAM statement

These steps are all required to properly state a KAM theorem and are, therefore, not truly part of the proof. The steps are:

- *Setting*: The precise context or setting of the KAM theorem. It includes the manifold, the local coordinates, and any optional structure defined on the manifold, which all vector fields and transformations must preserve. For well-known contexts, such as the Hamiltonian context, it is usually assumed that the reader is familiar with the setting, and it is common to see most of this step omitted.
- *Conditions*: These are the conditions required to prove the KAM theorem, such as the non-degeneracy and arithmetic conditions. This section can be quite extensive, depending on the chosen conditions.
- *Domain*: Here, we define the neighborhood of the invariant torus on which the KAM theorem is valid. This neighborhood must be explicitly defined to prove the theorem using the traditional KAM method.
- *Vector fields*: This step involves specifying the normal form of the integrable vector fields and perturbations, as well as properties such as differentiability and preservation of the structure. The normal form of the integrable vector field can differ significantly depending on the purpose of the KAM theorem.
- *KAM theorem statement*: With the stated preliminaries, it is possible to give the full statement of the KAM theorem.

KAM proof: Setup

This is the start of the KAM theorem proof. In this part, the iterative scheme is discussed, and the construction of Φ_∞ is fully specified. In this thesis, only Arnol'd's scheme is used, and this fact is reflected in the outline below.

- *Optional outline of the proof*: It is customary to give a brief outline of the KAM method central to KAM theory. This is optional and may be omitted for the sake of brevity, as usually no formal definitions are given[2]. As the outline is discussed here, it is omitted in the proof below.
- *Reduction to $\gamma = 1$ case*: This is a very common simplification due to the ease of reincorporating the γ parameter at the end of the proof.
- *Sequences and shrinking domains*: The sequences that govern the shrinking of the domains, the upper bound of the perturbation terms, and the ultraviolet cutoff are defined. These sequences must satisfy a range of conditions that are used in the rest of the proof, and it must be proven that such sequences do exist. With these sequences, the exact sequence of geometrically shrinking domains $(D_i)_i$ is given.
- *Iterative scheme*: The construction of Φ_i by the composition of the Ψ_i transformations, which are defined on the iterative domains D_i and proven to exist by the KAM step lemma. Under the assumption that Φ_i converges, which is proven later in the proof, some properties of the limit are proven.

[2]Note that omitting this step does not help with the understandability of a KAM proof.

KAM proof: KAM step lemma

In this part of the KAM theorem proof, we prove the existence of transformations Ψ_i and their properties. The existence of such Ψ_i is usually proven by a single lemma that inductively proves the existence of Ψ_i and all other required estimates and properties.

This part typically begins with three sections:

- *KAM step lemma statement*: The exact statement and conditions for the existence of suitable[3] Ψ_i.
- *Notation*: The introduction of plus and dot notation, which are common in KAM literature to prevent a flood of indices and constants in the proof of the KAM step lemma.
- *Intermediate domains*: To prove the required estimates and properties of each Φ_i, we make extensive use of Cauchy's estimate. This estimate requires a slightly shrunk domain each time it is applied, which is why we define several intermediate domains between D_{i+1} and D_i.

Each $\Psi_i : D_{i+1} \to D_i$ consists of two parts:

The structure-preserving transformation ψ_ω, which maps the (x, y) variables to their new values. These are constructed as the one-time flow map of a structure-preserving vector field V_ω, which is obtained by solving the linearized conjugation property.

The parameter shift ϕ, which maps the parameter to its new value in order to maintain the normal form of the new integrable part.

The proof of the KAM lemma constructs ψ_ω and ϕ separately and consists of the following six steps in this thesis:

- *Truncation*: The linearized conjugation property cannot be solved directly due to the small denominators, and must be approximated. To this end, the perturbation term is linearized and its Fourier series is truncated at the ultra-violet cut-off.
- *Solving the truncated linearized conjugation property*: The small denominators still present in the truncated equation are managed by an estimate, which allows us to obtain V_ω.
- *Coordinate transformation ψ_ω*: The transformation ψ_ω is defined by the one-time flow map of V_ω, and its well-definedness and properties are proven.
- *Parameter shift ϕ*: The parameter must be shifted to maintain the normal form of the new integrable part.
- *Full transformation Φ*: The full transformation Φ is constructed by the composition of ϕ_ω and ϕ.
- *New perturbation term*: Lastly, we prove that the norm of the new perturbation term is sufficiently small.

KAM proof: Convergence

The KAM step lemma is used to inductively construct the sequence of transformations $(\Psi_i)_i$, which in turn is used to construct the sequence of increasingly more accurate conjugations $(\Phi_i)_i$. To prove the KAM theorem, we must prove that this sequence uniformly converges to Φ_∞ on D_∞ and lastly extend it to the desired transformation Φ.

- *Estimates on Φ_i*: To help with the next steps, we make some estimates on each Φ_i.
- *Super-exponential decay*: To apply the Whitney inverse approximation lemma, we must first prove that $|\Phi_{i+1} - \Phi_i|_{D_{i+1}}$ decays super-exponentially. This is also the last step that puts additional conditions on the sequences.
- *Existence of the sequences*: We must prove that all of the conditions put on the sequences by the many estimates and lemmas can be satisfied simultaneously.
- *Whitney inverse approximation lemma*: With conveniently defined geometrically shrinking domains $(D_i)_i$ and the super-exponential decay proven we can apply the Whitney inverse approximation lemma to prove the existence of Φ_∞. Due to the uniform convergence Φ_∞ inherits most of the properties of Φ_i, most notably the linearity in the y-variable.
- *Whitney extension theorem*: We extend Φ_∞ to all of the frequency domain Σ by means of the Whitney extension theorem. This extended limit function is then used to construct the desired diffeomorphism Φ by means of exploiting the linearity in y-variable to extend Φ_∞ to all $y \in \mathbb{C}^n$. This finishes the proof for the $\gamma = 1$ case.
- *Reincorporation of γ*: The γ parameter is finally reincorporated by means of a simple scaling argument.

[3] With regards to the construction of Φ specified in the setup

4.2 KAM statement

In this section, we briefly review the setting of the KAM theorem and explicitly discuss the required preliminaries. These preliminaries include the context, the necessary conditions, the domain on which the diffeomorphism Φ and associated vector fields are defined, and the normal form of integrable systems.

Once these preliminaries have been established, we are able to provide the full formal statement of the parameterized KAM theorem.

4.2.1 Setting

This demonstration of KAM theory is set within the Lagrangian context, which is also known as the classical context for KAM theory.

Manifold

Let M be a connected $(2n)$-dimensional real manifold, i.e. $M \subset \mathbb{R}^{2n}$, on which a symplectic form ω_2 is defined. We assume that M contains a torus T^n, and that this torus is a Lagrangian submanifold of M.

Consequently, there exist local coordinates $(x \in T^n, y \in \mathbb{R}^n)$ such that M can be written as $M = T^n \times \mathbb{R}^n$, and under these coordinates, the symplectic form ω_2 takes the canonical form given by

$$\omega_2 = \sum_{l=1}^{n} dx_l \wedge dy_l. \tag{4.4}$$

The existence of such coordinates follows from Darboux's theorem and the assumption that T^n is a Lagrangian submanifold. We also define the maximum (or equivalently the supremum) norm on this manifold.

4.2.2 Conditions

In the parameterized KAM theorem the two conditions for KAM theory are somewhat simplified.

Non-degeneracy condition

Firstly we discuss the non-degeneracy condition, which is abstractly given by

Definition 4.2.1 (Non-degeneracy condition)**.** An integrable family of vector fields $\{X^\mu\}_\mu$ is *non-degenerate* at μ_0, if the frequency map $\omega(\mu) : P \to \mathbb{R}^n$ has a surjective derivative.

In this case, we have used the frequencies ω themselves as external parameters. This implies that the frequency map is the identity, and thus that the families $\{X^\omega\}_\omega$ are always non-degenerate.

Diophantine conditions

The Diophantine conditions are more implicitly used in frequency parameterized KAM theory. They are used to define the subset of Diophantine frequencies contained within the frequency (or parameter) domain $\Sigma \subset \mathbb{R}^n$.

The frequency domain Σ is usually assumed to be a small neighbourhood localized around a Diophantine ω_0, but the parameterized KAM theorem proven in this part only requires Σ to be compact. We define the Diophantine subset of the frequency domain Σ as

$$\Sigma_\gamma := \Sigma \cap \mathbb{R}^n_{\gamma,\tau}, \tag{4.5}$$

where $\mathbb{R}^n_{\gamma,\tau}$ is the set of Diophantine frequencies in $\mathbb{R}^n$, which is defined as

$$\mathbb{R}^n_{\gamma,\tau} := \left\{ \omega \in \mathbb{R}^n : |\langle k, \omega \rangle| \geq \frac{\gamma}{|k|^\tau} \quad \text{for all} \quad k \in \mathbb{Z}^n \right\}. \tag{4.6}$$

In this KAM theorem proof, we fix the constant τ at some value $\tau > n - 1$ for the entirety of the proof, but allow $\gamma > 0$ to vary. Lastly, we define a slightly shrunk subset

$$\Sigma'_\gamma := \{\omega \in \Sigma_\gamma : \operatorname{dist}(\omega, \partial\Sigma) \geq \gamma\} \tag{4.7}$$

of Σ_γ, which is required so that the parameter shift ϕ (defined on Σ'_γ) does not shift frequencies to values outside of Σ.

4.2.3 Domain

To formally define what a "small" perturbation P means in the context of KAM theory, we must first define a small compact domain containing the invariant tori $T^n \times 0 \times \Sigma$ over which the maximum norm can be taken. This domain should be large enough to contain all relevant dynamics about the invariant torus, but small enough for these dynamics to not differ too much from the parallel flow.

Furthermore, we must extend this domain to the complex numbers, as this is required for the application of the Paley-Wiener and Cauchy's estimates, both of which are extensively used in the KAM proof.

Let the domain D be a compact complex neighbourhood of $T^n \times \{0\} \times \Sigma$ in $(\mathbb{C}/2\pi\mathbb{Z})^n \times \mathbb{C}^n \times \mathbb{C}^n$ of the following form

$$D = \mathrm{cl}\left[U_\kappa(T^n) \times \mathcal{O}_\sigma \times U_\rho(\Sigma)\right], \tag{4.8}$$

where 'cl' stands for the closure and

$$\begin{aligned} U_\kappa(T^n) &:= \bigcup_{x \in T^n} \{x' \in (\mathbb{C}/2\pi\mathbb{Z})^n \,:\, |x' - x| < \kappa\}, \\ \mathcal{O}_\sigma &:= \{y \in \mathbb{C}^n \,:\, |y| < \sigma\}, \\ U_\rho(\Sigma) &:= \bigcup_{\omega \in \Sigma} \{\omega' \in \mathbb{C}^n \,:\, |\omega' - \omega| < \rho\}. \end{aligned}$$

Obviously, if the constants κ, σ, and ρ are sufficiently small, then the Hamiltonians N and P admit a holomorphic extension to D.

4.2.4 Vector fields

Firstly, let us consider the integrable vector fields. As shown in section 3.6, an n-parameter analytic family of integrable Hamiltonian vector fields $X = \{X^\mu\}_\mu$ on M is given by

$$X^\mu = \omega(\mu)\frac{\partial}{\partial x}. \tag{4.9}$$

And, as mentioned in the introduction of this part, we parameterize the integrable vector fields by the frequencies $\omega \in \mathbb{R}^n$. Thus, we have $X = \{X^\omega\}_\omega$ which is given by

$$X^\omega = \omega\frac{\partial}{\partial x}. \tag{4.10}$$

Associated with X is the parameterized Hamiltonian N, given by

$$N(x, y, \omega) = \langle \omega,\, y\rangle + e(\omega), \tag{4.11}$$

where $e(\omega)$ is constant for all (x, y) and thus of little dynamical meaning[4]. Using Hamiltonian functions simplifies the calculations and estimates in KAM theory. Therefore, for the rest of this part, we only consider vector fields in terms of their Hamiltonian functions.

The perturbations of X are considered in terms of their Hamiltonian function P, which are relatively simple to define. Let $P : D \to \mathbb{R}$ be an analytic function such that $|P|_D < \delta$, where $\delta > 0$ and $|\cdot|_D := \sup_{x \in D} |P(x)|$ is the maximum norm over the (compact) domain D.

4.2.5 Formal statement

With all preliminaries out of the way, we can formally state the parameterized KAM theorem.

Theorem 4.2.2 (Parameterized KAM theorem.)**.** *Let $H = N + P$ be an analytic perturbation of an integrable Hamiltonian* (4.11) *defined on D. Then, there exists a constant $\delta > 0$ such that if*

$$|P|_D < \delta, \tag{4.12}$$

then there exists a Whitney regular diffeomorphism $\Phi : U_\kappa(T^n) \times \mathbb{C}^n \times \Sigma \to U_\kappa(T^n) \times \mathbb{C}^n \times \Sigma$ such that:

- *Φ is holomorphic and symplectic for fixed ω.*
- *Φ is C^∞-close to the identity.*
- *On $U_\kappa(T^n) \times \mathbb{C}^n \times \Sigma'_\gamma \cap \Phi^{-1}(U_\kappa(T^n) \times \{0\} \times \Sigma)$, the Hamiltonian vector field corresponding to $H \circ \Phi = N_\infty$ is integrable (or equivalently in the form* (4.11)*).*

Theorem 4.0.2 is a direct corollary of the above theorem.

[4] It is common to see the constant terms omitted in the entire KAM proof.

4.3 KAM proof: Setup

In this section, we set up the KAM proof.

We prove theorem 4.2.2 using the traditional KAM method with Arnol'd's iteration scheme. Arnol'd's scheme allows us to prove the persistence of the invariant tori for all frequencies in Σ'_γ, instead of just a single Diophantine frequency ω_0. We construct a sequence $(\Phi_i)_i$ of increasingly more accurate approximate conjugations on increasingly shrinking domains $(D_i)_i$. The limit of these sequences are Φ_∞ and $D_\infty = U_\kappa(T^n) \times 0 \times \Sigma'_\gamma$ respectively, where Φ_∞ conjugates the invariant tori present within D_∞.

Before delving into the iteration scheme, we first discuss the sequences used for the construction of the shrinking domains $(D_i)_i$, the bound of the new perturbation after applying the approximate conjugation Φ_i, and the ultraviolet cut-off necessary to manage the small denominators. We then use these sequences to define the shrinking domains, and use these domains to formally define the construction of $(\Phi_i)_i$ by means of coordinate transformations $(\Psi_i)_i$ and discuss some of their properties.

Reduction to the $\gamma = 1$ case

It is sufficient to prove theorem 4.2.2 for only the case $\gamma = 1$. Different cases with differing values of γ can be easily transformed to the $\gamma = 1$ case using a rescaling argument. We can rescale the time from t to γt and the parameter ω to $\gamma^{-1}\omega$, which changes the frequency domain to $\gamma^{-1}\Sigma$. After rescaling, the new frequency domain is unbounded as γ approaches zero, and thus we must consider this in the proof.

In the next sections, we assume that $\gamma = 1$, which allows us to set $\rho = 1$ by extension. The actual value of γ is reintroduced at the end of the proof.

Dot notation

To prevent a flood of constants we introduce the dot notation. We write $a \lessdot b$ and $a \cdot\!< b$ if there exists an explicit positive constant $c \geq 1$, which depends only on τ and n, such that $a \leq cb$ or $ca \leq b$, respectively.

4.3.1 Sequences and shrinking domains

In this section, we first define the sequences and afterwards explain what they are used for. We define the geometric and exponential sequences as follows:

$$s_i = \tfrac{1}{4} s_{i-1} = 4^{-i} s_0, \tag{4.13}$$

$$r_i = \tfrac{1}{2} s_i^{2\tau+2}, \tag{4.14}$$

$$d_i = [s_i^{-2}], \tag{4.15}$$

$$\delta_i = \delta_{i-1}^{\chi/2} \delta_{i-1} = \delta_0^{(1+\chi/2)^i}, \tag{4.16}$$

$$\varepsilon_i = \delta_i^{1/2} = \varepsilon_{i-1}^{\chi/2} \varepsilon_{i-1}, \tag{4.17}$$

with $0 < \chi < 1$. The justification for these definitions is briefly mentioned below and covered in more detail in the associated estimates in appendix A.

The shrinking domains

The sequences $(s_i)_i$, $(\varepsilon_i)_i$, and $(r_i)_i$ are used to define the shrinking domains

$$D_i = \mathrm{cl}\left[U_{\kappa+s_i}(T^n) \times \mathcal{O}_{\varepsilon_i} \times U_{r_i}(\Sigma'_1)\right], \tag{4.18}$$

where cl stands for the closure. These shrinking domains serve three purposes:

- On these increasingly smaller domains, an increasingly sharper small denominator estimate holds. This lesser condition is good enough to ensure a solution to the truncated linearized conjugation relation. This condition eventually becomes equivalent to the Diophantine conditions as i goes to infinity.
- On these shrinking domains, the domain of the y variables shrinks fast enough to allow us to prove a sufficient estimate on the norm of each new perturbation term. This is required to prove that the perturbation term approaches zero as i goes to infinity.
- Lastly, the shrinking domains are constructed specifically to allow for the application of the Whitney inverse approximation lemma, which proves the convergence of $(\Phi_i)_i$ and the smoothness of the limit in the Whitney sense.

Lastly, note that the limit of these domains approaches the domain

$$\lim_{i\to\infty} D_i = D_\infty = U_\kappa(T^n) \times \{0\} \times \Sigma'_1. \tag{4.19}$$

The bounds for the perturbations

The exponentially shrinking sequence $(\delta_i)_i$ is used to dominate the norm of each new perturbation P_i on D_i at each step i.

This sequence, together with $(\varepsilon_i)_i$, could have been chosen to shrink a lot slower[5] and the new perturbation would still approach zero. However, in order to prove the Whitney regularity of Φ_∞ in the ω parameter, we must also show that $|\Phi_{i+1} - \Phi_i|_{D_i}$ decays exponentially fast, which does require $(\delta_i)_i$ to shrink exponentially fast.

The ultraviolet cut-off of the Fourier series

The geometrically growing integer sequence $(d_i)_i$ serves as the ultra-violet cut-off for the Fourier truncation of the perturbation P_i at each step.

This cut-off is used in the estimate for the lower bound of the small denominators $|i\langle k, \omega\rangle|$ for $0 < |k| < d_i$ on each domain D_i. As i approaches infinity, we have that d_i approaches infinity, and that the small denominator estimate approaches the true Diophantine conditions.

Requirements on the sequences

Firstly, to ensure that the sequences $(s_i)_i$ and $(\delta_i)_i$ do in fact shrink and the domains D_i are well-defined, we require

$$0 < s_0 < \min\{\kappa, \tfrac{1}{2}\} \quad \text{and} \quad 0 < \delta_0 < \tfrac{1}{2}. \tag{4.20}$$

Additionally, the estimates used to prove the KAM step lemma and the convergence of the approximate conjugation require the sequences to satisfy certain conditions. These conditions are taken into account in the choice of the values of s_0 and δ_0. Therefore, s_0 and δ_0 need to satisfy various inequalities of the form:

$$s_0 \leq c, \quad \delta_0 \leq c, \quad \delta_0^\zeta \leq c s_0 \quad \text{and} \quad s_0 < \frac{1}{c - \zeta \ln \delta_0}. \tag{4.21}$$

with multiple constants $c > 0$ and exponents $\zeta > 0$, which both only depend on n, τ and the definition the sequences. These constants and exponents remain constant[6] throughout the KAM proof, and each c and ζ can be made explicit.

The fourth inequality is only used once, and is required to estimate the higher order Fourier coefficients of the perturbations P_i. Unlike the other conditions, it restricts s_0 as a function of δ_0, which is the "opposite" direction. Thus, it must be explicitly verified that it can be satisfied simultaneously with the other required conditions.

All inequalities are made explicit in the KAM step lemma and convergence part of the KAM proof, and it is proven in lemma 4.5.4 that such s_0 and δ_0 do in fact exist.

4.3.2 Iterative scheme

We aim to construct a Whitney regular diffeomorphism

$$\Phi : U_\kappa(T^n) \times \mathbb{C}^n \times \Sigma \to U_\kappa(T^n) \times \mathbb{C}^n \times \Sigma, \tag{4.22}$$

such that, when Φ is restricted to $U_\kappa(T^n) \times \mathbb{C}^n \times \Sigma'_\gamma \cap \Phi^{-1}(U_\kappa(T^n) \times \{0\} \times \Sigma)$, we have $H_0 \circ \Phi = N_\infty$, where N_∞ is integrable or equivalently in the form (4.11).

The KAM method constructs a conjugation Φ_∞ on the restricted domain $D_\infty := U_\kappa(T^n) \times \{0\} \times \Sigma'_\gamma$ by iteratively approaching the desired conjugation with successively better approximations. The conjugation Φ_∞ is then extended to the entire domain $U_\kappa(T^n) \times \mathbb{C}^n \times \Sigma$ to obtain Φ. This extension is covered in more detail in section 4.5.

[5] While still decaying sufficiently fast in relation to $(s_i)_i$.

[6] Hence, the use of the term "constant".

Construction of Φ_∞

We construct Φ_∞ as the limit of the sequence of successively better approximations $(\Phi_i)_i$, which is given by

$$\Phi_i : D_i \to \Phi_i(D_i) \subset D_0 \tag{4.23}$$

$$\Phi_i := \Psi_0 \circ \Psi_1 \circ \cdots \circ \Psi_{i-1}, \tag{4.24}$$

where each transformation Ψ_i is defined by

$$\Psi_i : D_{i+1} \to \Psi_i(D_{i+1}) \subset D_i, \tag{4.25}$$

such that

$$H_i \circ \Psi_i = H_{i+1} = N_{i+1} + P_{i+1}, \tag{4.26}$$

where N_{i+1} is integrable and in the normal form (4.11) and the perturbation P_{i+1} satisfies $|P_{i+1}|_{D_{i+1}} < \delta_{i+1}$. Additionally, each Ψ_i is required to be symplectic and holomorphic for fixed ω, and C^∞-close to the identity map. The existence of each transformation Ψ_i is proven inductively by the KAM step lemma 4.4.1.

Properties of each Φ_i

We consider the properties of each successive approximation Φ_i.

Firstly, by using (4.24) and (4.25), we can see that

$$H \circ \Phi_{i+1} = H \circ \Phi_i \circ \Psi_i = H_i \circ \Psi_i = H_{i+1} = N_{i+1} + P_{i+1}. \tag{4.27}$$

Secondly, since for fixed ω every transformation Ψ_i is symplectic and holomorphic, we have that every transformation Φ_i is also symplectic and holomorphic. Similarly, we can conclude that Φ_i is C^∞-close to the identity map.

Properties of Φ_∞

The sequence $(\Phi_i)_i$ converges uniformly on the domain

$$D_\infty = \bigcap_i D_i = U_\kappa(T^n) \times \{0\} \times \Sigma_1' \tag{4.28}$$

to the transformation $\Phi_\infty := \lim_{i\to\infty} \Phi_{i+1}$. The uniform convergence is proven explicitly in section 4.5, by means of the Whitney inverse approximation lemma.

Due to this uniform convergence, the transformation Φ_∞ inherits the properties of Φ_i. Thus, we have that Φ_∞ is symplectic and holomorphic for fixed ω, and that Φ_∞ is C^∞-close to the identity map.

Moreover, we have $H \circ \Phi_\infty = N_\infty + P_\infty$ on D_∞, where

$$N_\infty := \lim_{i\to\infty} N_i \quad \text{and} \quad P_\infty := \lim_{i\to\infty} P_i. \tag{4.29}$$

Since each N_{i+1} is integrable and in normal form at each step, we have that N_∞ is also integrable and in normal form. Additionally, since each P_{i+1} satisfies $|P_{i+1}| < \delta_{i+1}$, we have that

$$|P_\infty| = \left|\lim_{i\to\infty} P_{i+1}\right| \leq \lim_{i\to\infty} \delta_i = 0. \tag{4.30}$$

Therefore, the perturbation P_∞ is in fact zero and $H \circ \Phi_\infty = N_\infty$. Thus, Φ_∞ conjugates the perturbed system H to an integrable system N_∞ on D_∞.

What remains to be done is to explicitly prove the existence of each transformation Ψ_i, the uniform convergence of $(\Phi_i)_i$ to Φ_∞ on D_∞, and finally prove that the extension of Φ_∞ to D, denoted by Φ, satisfies all properties stated in theorem 4.2.2.

4.4 KAM proof: KAM step lemma

This section is entirely dedicated to the KAM step lemma, which inductively proves the existence of the transformations Ψ_i, as described in section 4.3.

Lemma 4.4.1 (KAM step lemma)**.** *For a given $H_i = N_i + P_i$, there exists a coordinate transformation*

$$\begin{array}{rccc} \Psi_i : & D_{i+1} & \to & D_i \\ & (x, y, \omega_+) & \mapsto & (\psi_\omega(x, y), \omega), \end{array} \tag{4.31}$$

where ψ_ω is symplectomorphism for fixed ω, and such that $H_i \circ \Psi_i = H_{i+1}$ has the form

$$H_{i+1} = N_{i+1} + P_{i+1}, \tag{4.32}$$

where N_{i+1} has the same form as (4.11) *and the perturbation P_{i+1} satisfies*

$$|P_{i+1}|_{D_{i+1}} < \delta_{i+1}. \tag{4.33}$$

In the proof of the KAM step lemma, we use the methods from the previous two parts to construct a suitable transformation Ψ_i. After constructing the transformation, we examine the resulting new integrable part N_{i+1} and perturbation P_{i+1}, and prove their respective properties. The KAM step lemma is usually the largest part of any KAM theorem proof, and within the proof of the lemma, we manage the small denominators and the requirements imposed by the geometry of the manifold[7].

Notation

We introduce additional notation to help with the readability of the proof.

Firstly, since the KAM lemma only concerns itself with a single iteration step, only indices $i+1$ and i are used. We simplify this by using the plus notation, which replaces the index $i+1$ by $+$ and entirely omits the index i. Secondly, we use the shorthand $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Lastly, the multi-index notation is extensively used in the proof of the lemma and as well as in appendix A, where the estimates are explicitly proven.

Intermediate domains

Within the proof of the KAM step lemma, we introduce the intermediate domains D^ι between D_+ and D. These intermediate domains are necessary to prove the estimates that use Cauchy's inequality, which requires the domain to be slightly shrunk at each application. Recall that the relations between s_+, r_+, ε_+ ands_i, r_i, and ε_i are given by

$$s_+ = \frac{1}{4}s, \qquad r_+ = 4^{-2\tau-2}r, \qquad \varepsilon_+ = \varepsilon^{\chi/2}\varepsilon. \tag{4.34}$$

With this knowledge, we can define

$$s_\iota = s - \frac{\iota}{8}s, \qquad r_\iota = r - \frac{\iota}{8}r, \qquad \varepsilon_\iota = 2^{-\iota}\varepsilon \quad \text{with} \quad \iota \in \{1, \cdots 4\}, \tag{4.35}$$

and

$$s^*_\iota = s - \frac{4+\iota}{8}s, \qquad r^*_\iota = r - \frac{4+\iota}{8}r, \qquad \varepsilon^*_\iota = 2^{4-\iota}\varepsilon_+ \quad \text{with} \quad \iota \in \{1, \cdots 3\}, \tag{4.36}$$

to define two sets of intermediate domains

$$D^\iota := U_{\kappa+s_\iota}(T^n) \times \mathcal{O}_{\varepsilon_\iota} \times U_{r_\iota}(\Sigma_1') \quad \text{and} \quad D^\iota_* := U_{\kappa+s^*_\iota}(T^n) \times \mathcal{O}_{\varepsilon^*_\iota} \times U_{r^*_\iota}(\Sigma_1'), \tag{4.37}$$

which results in seven intermediate domains satisfying

$$D_+ \subset D^3_* \subset \cdots \subset D^1_* \subset D^4 \subset \cdots \subset D^1 \subset D \tag{4.38}$$

Additionally, in order for this to hold true, we must make the assumption that $\varepsilon^{\chi/2} < \frac{1}{128}$ whence

$$8\varepsilon_+ = 8\varepsilon^{\chi/2}\varepsilon < \tfrac{1}{16}\varepsilon. \tag{4.39}$$

We require two sets of intermediate domains. The lesser set, denoted by D^ι_*, is used to estimate the norm of the truncation $P_{\text{lin},d}$. The greater set, denoted D^ι, is used to estimate the norm of the Hamiltonian F and its derivatives. For more details, see appendix A.

[7] In this case, that Ψ is a symplectomorphism.

4.4.1 KAM step lemma proof

Each transformation Φ consists of two parts: the transformation ψ_ω and the parameter shift $\phi : \omega_+ \mapsto \omega$.

We start by constructing ψ_ω as the time-one flow map of a Hamiltonian vector field V_ω with the Hamiltonian function F, which is determined by the equations described in section 3.7.

The Hamiltonian F is obtained by solving the linearized conjugation relation

$$\hat{N}_+ = N + \{N, F\} + P_{\mathrm{lin},d}, \tag{4.40}$$

where we have set the new integrable part to

$$\hat{N}_+ = N + [P_{\mathrm{lin},d}(\cdot, y, \omega)] = N + \frac{1}{\mathrm{Vol}(T^n)} \int_{T^n} P_{\mathrm{lin},d}(x, y, \omega)\, dx. \tag{4.41}$$

We solve this equation step by step and give the necessary estimates at each step.

Linearization and Fourier truncation of P

Before solving (4.40) for F, we first need to calculate $P_{\mathrm{lin},d}$ and $[P_{\mathrm{lin},d}(\cdot, y, \omega)]$. In this step, we linearize and truncate the perturbation P, derive $[P_{\mathrm{lin},d}(\cdot, y, \omega)]$, and provide the estimates on $P_{\mathrm{lin},d}$ and $P - P_{\mathrm{lin},d}$.

To begin, we consider the Taylor series of the perturbation P in the y variables about the origin

$$P(x, y, \omega) = \sum_{|j| \in \mathbb{N}_0^n} P_j(x, \omega) y^j, \tag{4.42}$$

and the Fourier series of each coefficient function P_j in x

$$P_j(x, \omega) = \sum_{k \in \mathbb{Z}^n} P_j^k(\omega) e^{2\pi \langle k, x \rangle}. \tag{4.43}$$

We linearize[8] P in the y variable and truncate the Fourier series of the linearized P at the ultraviolet cut-off d to obtain

$$P_{\mathrm{lin},d} = \sum_{|j| \leq 1} \sum_{|k| < d} P_j^k(\omega) e^{2\pi \langle k, x \rangle} y^j. \tag{4.44}$$

The average of $P_{\mathrm{lin},d}$ over x is determined by the zeroth Fourier coefficients. Therefore, we have

$$[P_{\mathrm{lin},d}(\cdot, y, \omega)] = P_0^0(\omega) + \sum_{|j|=1} P_j^0(\omega) y^j = P_0^0(\omega) + P_1^0(\omega). \tag{4.45}$$

It is important to note that the linearization and Fourier truncation of a Hamiltonian function preserve the symplectic structure. This property is evident in the Hamiltonian context but must be verified for other contexts to ensure that these operations preserve the structure-preserving property of the perturbation. Therefore, it is one of the requirements for an admissible structure.

We also define $\hat{P} = P - P_{\mathrm{lin},d}$, which is the remainder of the higher-order terms in y variables and the higher-frequency Fourier terms. Both the linearization and truncation have sufficiently small remainders on the domain D for the ultraviolet cut-off d. This is proven by the following estimate.

Lemma 4.4.2 (Truncation estimates)**.** *Under the assumptions that*

$$d^n e^{-\frac{5}{8} ds} < \varepsilon, \tag{4.46}$$

$$ds > \frac{8}{5} n, \tag{4.47}$$

the truncation $P_{lin,d}$, as given in (4.44)*, satisfies*

$$|P_{lin,d}|_{D^1} \lessdot \delta, \tag{4.48}$$

$$|\hat{P}|_{D^1_*} = |P - P_{lin,d}|_{D^1_*} \lessdot \varepsilon^\chi \delta. \tag{4.49}$$

This lemma is proven by estimating the remainder of the Taylor series in the y variable using Cauchy's estimate, and estimating the high-frequency part of the Fourier series using a suitable Paley-Wiener estimate. For a complete proof of this lemma, please refer to the appendix (lemma A.2.1).

[8] It is possible to truncate in the y variable at an arbitrarily high order, but this is not necessary for this proof.

Solving equation (4.40)

With $P_{\text{lin},d}$ and $[P_{\text{lin},d}(\cdot, y, \omega)]$ calculated, we can solve the linearized conjugation relation (4.40) for F to obtain

$$F(x, y, \omega) = \sum_{|j|\leq 1} \sum_{0<|k|\leq d} F_k^j(\omega) e^{i\langle k,\, x\rangle} y^j \quad \text{with} \quad F_k^j(\omega) = \frac{P_k^j(\omega)}{i\langle k,\, \omega\rangle}. \tag{4.50}$$

For a complete derivation, please see appendix A.1.

We see that the solution to F contains the small denominators $i\langle k, \omega\rangle$. We estimate these small denominators up to the ultraviolet cut-off d, as was done in the demonstration of the one-bite problem (cf. lemma 2.4.4).

Lemma 4.4.3 (Small denominator estimate)**.** *For all $\omega \in U_r(\Sigma_1')$ and $k \in \mathbb{Z}^n$ with $0 < |k| \leq d$ one has*

$$|\langle \omega, k\rangle| \geq |k|^{-\tau}. \tag{4.51}$$

The small denominator estimate is the result of cleverly chosen r and d. For a complete proof of this lemma, refer to the appendix (lemma A.2.2).

The small denominator estimate manages the small denominators in the coefficients of the Fourier series of F, thus ensuring that F is well-defined. It is also used directly to prove the following estimates.

Lemma 4.4.4 (Estimates on F)**.** *The Hamiltonian function F, as given in* (4.50)*, satisfies*

$$\begin{array}{cccc} s^{2\tau+1}|F|_{D^2}, & s^{2\tau+2}\,|\partial_x F|_{D^3}\,, & \varepsilon s^{2\tau+1}\,|\partial_y F|_{D^3} & <\!\cdot \quad \delta, \\ s^{2\tau+3}\,|\partial_{xx} F|_{D^4}\,, & \varepsilon s^{2\tau+2}\,|\partial_{xy} F|_{D^4}\,, & \varepsilon s^{2\tau+2}\,|\partial_{yx} F|_{D^4} & <\!\cdot \quad \delta. \\ & r s^{2\tau+2}\,|\partial_{x\omega} F|_{D^4}\,, & r\varepsilon s^{2\tau+2}\,|\partial_{y\omega} F|_{D^4} & <\!\cdot \quad \delta. \end{array} \tag{4.52}$$

and lastly $\partial_{yy} F = 0$.

For a proof of this lemma consult the appendix (lemma A.2.3).

Coordinate transformation ψ_ω

With a solution for the Hamiltonian function F, we can define the Hamiltonian vector field V_ω. Let us fix $\omega \in U_r(\Sigma_1')$ to ensure a well-defined F, and let $F_\omega(x, y) := F(x, y, \omega)$. The vector field V_ω is then given by

$$V_\omega = \partial_y F_\omega \frac{\partial}{\partial x} - \partial_x F_\omega \frac{\partial}{\partial y}. \tag{4.53}$$

Let $\varphi_{V_\omega}^t$ be the flow of V_ω and define $\psi_\omega := \varphi_{V_\omega}^1$. This coordinate transformation is then used to define

$$\psi : D_*^2 \to D_*^1 : (x, y, \omega) \mapsto (\psi_\omega(x, y), \omega). \tag{4.54}$$

As ψ_ω is constructed from the flow of the analytic Hamiltonian vector field V_ω, we have that ψ_ω, and in extension ψ, are a symplectomorphisms. The well-definedness of ψ and the corresponding estimates are proven by the following lemma.

Lemma 4.4.5. *Under the assumption that*

$$\delta \cdot\!< \varepsilon_+ s^{2\tau+2} < \varepsilon s^{2\tau+2}, \tag{4.55}$$

we have that the transformation ψ, as given in (4.54)*, is well-defined and satisfies*

$$s^{2\tau+2}|\psi_\omega - id|_{D_*^2} <\!\cdot\, \varepsilon^{-1}\delta, \tag{4.56}$$

$$r s^{2\tau+2}|D\psi_\omega - Id|_{D_*^2} <\!\cdot\, \varepsilon^{-1}\delta, \tag{4.57}$$

where id is the identity map and Id is the identity matrix.

For a full proof of this lemma consult the appendix (lemma A.2.4).

Lastly, it is important to note that since the perturbation P_{lin} is linear in the y-variable, so are the Hamiltonian functions F_ω by construction (see (4.50)). This linearity extends to the vector fields V_ω generated by the Hamiltonian functions F_ω, and the time-one flow maps ψ_ω induced by these vector fields on M.

Parameter shift ϕ

We continue with the construction of the parameter shift $\phi : \omega_+ \mapsto \omega$. The goal of the parameter shift is to put the new integrable part $\hat{N}_+$ back into the normal form (4.11). Recall that we have chosen the following $\hat{N}_+$,

$$\hat{N}_+ = N + [P_{\text{lin},d}(\cdot, y, \omega)] = \langle \omega, y \rangle + e(\omega) + P_0^0(\omega) + \sum_{|j|=1} P_j^0(\omega) y^j, \tag{4.58}$$

where $P_0^0(\omega)$ is constant for all (x, y). Thus, it does not impact the parameter ω.
We also substitute

$$\sum_{|j|=1} P_j^0(\omega) y^j = \sum_{i=0}^{n} P_{j(i)}^0(\omega) \cdot y_i = \langle \Lambda(\omega), y \rangle, \tag{4.59}$$

with $j(i) = (0, \cdots, 0, \overset{i}{1}, 0, \cdots, 0)$, where y_i is the i-th y variable, and lastly where $\Lambda(\omega)$ is the vector $\left(P_{j(0)}^0, P_{j(1)}^0, \cdots, P_{j(n)}^0\right) \in \mathbb{R}^n$.
Thus, we obtain

$$\hat{N}_+ = \langle \omega + \Lambda(\omega), y \rangle + e(\omega) + P_0^0(\omega). \tag{4.60}$$

For N_+ to be in the normal form (4.11), we must construct a parameter shift ϕ such that

$$\hat{N}_+ \circ \phi(x, y, \omega_+) = N_+(x, y, \omega_+) = \langle \omega_+, y \rangle + e_+(\omega_+). \tag{4.61}$$

Since Λ is small due to it being derived from a small perturbation P, we have that the map

$$\sigma : \omega \mapsto \omega + \Lambda(\omega), \tag{4.62}$$

is close to the identity and thus invertible by the implicit function theorem. We use this inverse to define

$$\phi : D_*^3 \to D_*^2 : (x, y, \omega_+) \mapsto (x, y, \sigma^{-1}(\omega_+)) = (x, y, \omega). \tag{4.63}$$

It is easily verified that ϕ indeed satisfies (4.61), which gives us the new integrable part in normal form N_+ with the new frequency defined by $\omega_+ = \sigma(\omega)$, and the new constant term given by $e_+(\omega_+) = e(\sigma^{-1}(\omega_+)) + P_0^0(\sigma^{-1}(\omega_+))$.
The well-definedness of ϕ and the corresponding estimates are proven by the following lemma.

Lemma 4.4.6 (Parameter shift ϕ lemma)**.** *Under the assumption that*

$$\delta \cdot\!< \varepsilon r = \varepsilon s^{2\tau+2}, \tag{4.64}$$

we have that the parameter shift ϕ, as given in (4.63)*, is well-defined and satisfies*

$$|\phi - id|_{D_*^3} \lessdot \varepsilon^{-1}\delta, \tag{4.65}$$

$$r|D\phi - Id|_{D_*^3} \lessdot \varepsilon^{-1}\delta, \tag{4.66}$$

where id is the identity map, and Id is the identity matrix.

For a proof of this lemma consult the appendix (lemma A.2.5).

Again it is important to note that the parameter shift ϕ is (technically) linear in the y-variable.

Full transformation Ψ

With the coordinate transformation ψ and the parameter shift ϕ, we can finally define the full transformation Ψ, which is given by

$$\Psi : \Psi := \psi \circ \phi : D^3_* \to D^1_* : (x, y, \omega_+) \mapsto (\psi_\omega(x, y), \omega). \tag{4.67}$$

This map is well-defined, as both ψ and ϕ have been shown to be well-defined, and their corresponding estimates are provided in the following lemma.

Lemma 4.4.7 (Estimates on Ψ)**.** *The coordinate transformation Ψ, as given in* (4.67)*, is well-defined and satisfies*

$$s^{2\tau+2}|\Psi - id|_{D^3_*} <\cdot \varepsilon^{-1}\delta \tag{4.68}$$

$$rs^{2\tau+2}|D\Psi - Id|_{D^3_*} <\cdot \varepsilon^{-1}\delta, \tag{4.69}$$

where id is the identity map, and Id is the identity matrix.

This lemma is proven by the use the chain rule, lemma 4.4.5 and lemma 4.4.6. For a complete proof of this lemma consult the appendix (lemma A.2.6).

Additionally, since both ψ and ϕ are linear in the y-variable, so is their composition Ψ.

With lemma 4.4.7, we are very close to proving the KAM step lemma.

- Firstly, since $D_+ \subset D^3_*$ and $D^1_* \subset D$, lemma 4.4.7 implies that the restriction of Ψ to D_+ is well-defined and maps into D.
- Secondly, the construction of ψ_ω implies that it is a symplectomorphism for fixed ω.
- Thirdly, lemma 4.4.7 implies that Φ is close to the identity.
- Lastly, the construction of Φ implies that $H \circ \Phi = N_+ + P_+$, where N_+ has the form (4.11).

There is only one step left to prove lemma 4.4.1, we need to verify that the new perturbation term P_+ satisfies $|P_+|D+ <\cdot \delta_+$.

New perturbation term

To obtain the new perturbation term P_+, we apply the transformation Ψ to H to obtain

$$H \circ \Psi = H \circ \psi_\omega \circ \phi = N_+ + P_+, \tag{4.70}$$

where P_+ is given by

$$P_+ = \left(\int_0^1 \{(1-t)\{N, F\} + tP_{\text{lin},d}, F\} \circ \psi_\omega \, dt \right) \circ \phi + (P - P_{\text{lin},d}) \circ \psi_\omega \circ \phi. \tag{4.71}$$

For the derivation of the integral term, please refer to appendix A.1. All the estimates made so far culminate in the following lemma.

Lemma 4.4.8. *Under the assumption that*

$$\delta \cdot< \varepsilon^{1+\chi} s^{2\tau+2}, \tag{4.72}$$

the new perturbation term P_+ satisfies

$$|P_+|_{D_+} \leq \delta_+. \tag{4.73}$$

For a full proof of this lemma consult the appendix (lemma A.2.7).

The proof of the KAM step lemma is completed by lemma 4.4.8 under the assumption that conditions (4.46), (4.47), (4.55), (4.64), and (4.72) hold for the sequences. Lemma 4.5.4 proves the existence of such sequences.

4.5 KAM proof: Convergence

There are several steps remaining in the proof of theorem (4.2.2). These are discussed in this section.

Firstly, we need to prove that the sequence of approximate conjugations $(\Phi_i)_i$ satisfies the super-exponential[9] decay condition of the Whitney inverse approximation lemma. To do this, we must show that the C^1-norm of Φ_i on D_i, given by $||\Phi_i||_{1,D_i} := \max(|\Phi|_{D_i}, |D\Phi_i|_{D_i})$, is universally bounded for all $i \in \mathbb{N}_0$.

Secondly, we use the Whitney inverse approximation lemma to obtain the limit function Φ_∞ and use the Whitney extension theorem to extend this limit function to the complete frequency domain. We can then construct the desired transformation Φ from the extended limit function.

Lastly, we reintroduce the γ parameter and prove that all assumptions on the sequences made in section 4.4 and in this section can be satisfied by choosing a suitable s_0 and δ_0. This completes the proof of theorem 4.2.2.

4.5.1 Estimates on Φ_i

To begin, we construct the sequence $(\Psi_i)_i$ inductively by applying the KAM step lemma to H_i. We start with the Hamiltonian $H_0 = H$, where H is the original perturbed Hamiltonian given in theorem 4.2.2. With the sequence $(\Psi_i)_i$, we easily obtain the sequence $(\Phi_i)_i$.

We must estimate the C^1-norm of each Φ_i on D_i, which is done in the following lemma.

Lemma 4.5.1 (Estimates of Φ_{i+1})**.** *The transformation Φ_{i+1} satisfies*

$$||\Phi_{i+1}||_{1,D_{i+1}} \lesssim ||\Phi_i||_{1,D_i} \cdot \max(|\Psi_i - id|, |D\Psi_i - Id|) \tag{4.74}$$

$$|\Phi_{i+1} - \Phi_i|_{D_{i+1}} \lesssim ||\Phi_i||_{1,D_i} \cdot |\Psi_i - id| \tag{4.75}$$

for all $i \in \mathbb{N}_0$.

For a proof of this lemma consult the appendix (lemma A.3.1). This lemma can be directly applied to prove that $||\Phi_i||_{1,D_i}$ is universally bounded for all $i \in \mathbb{N}_0$.

Lemma 4.5.2 (Universal bound of $||\Phi_i||_{1,D_i}$)**.** *Under the assumption that*

$$\delta_0^{\chi/4} < 4^{-4\tau-4} \tag{4.76}$$

the sequence $(||\Phi_i||_{1,D_i})_i$ is universally bounded, i.e. there exists a constant $C > 0$ such that

$$||\Phi_i||_{1,D_i} \leq C \quad \textit{for all} \quad i \in \mathbb{N}_0. \tag{4.77}$$

For a proof of this lemma consult the appendix (lemma A.3.2).

Super-exponential decay of $|\Phi_{i+1} - \Phi_i|_{D_{i+1}}$

With lemma 4.5.1 and 4.5.2, we can prove that $|\Phi_{i+1} - \Phi_i|_{D_{i+1}}$ super-exponentially decays for $0 \leq k \leq 1$.

The super-exponential decay of $|\Phi_{i+1} - \Phi_i|_{D_{i+1}}$ allows us to apply the Whitney approximation lemma to prove that the sequence $(\Phi_i)_i$ is uniformly convergent to the limit function Φ_∞ on $D_\infty := \mathrm{cl}\,[U_\kappa(T^n)] \times \{0\} \times \Sigma_1'$. Moreover, the lemma also proves that Φ_∞ is Whitney-smooth in the ω parameter.

Lemma 4.5.3 (Super-exponential decay of $|\Phi_{i+1} - \Phi_i|_{D_{i+1}}$)**.** *Under the assumption that for an arbitrary $0 < b < \frac{1}{2}$ we have*

$$\delta_0^{1/2-b} < s_0^{2\tau+2} \quad \textit{and} \quad \delta_0^{\chi/2(1/2-b)} 4^{2\tau+2} < 1, \tag{4.78}$$

there exists a constant $M > 0$ such that

$$|\Phi_{i+1} - \Phi_i|_{D_{i+1}} < M\delta_0^{b^i} \tag{4.79}$$

i.e. the norm $|\Phi_{i+1} - \Phi_i|$ on D_{i+1} converges super-exponentially fast to zero.

For a proof of this lemma consult the appendix (lemma A.3.3). This is the last estimate of the KAM proof and it introduces our last assumption.

[9] Geometric decay is also sufficient.

4.5.2 Whitney inverse approximation lemma

In this subsection, we apply the Whitney inverse approximation lemma (cf. lemma 2.3.3) to a sequences of coefficient functions $(\Phi_i)_i$. This requires a geometrically shrinking sequence of domains D_i on which $|\Phi_{i+1} - \Phi_i|_{D_{i+1}}$ decays super-exponentially.

The super-exponential decay of $|\Phi_{i+1} - \Phi_i|_{\bar{D}_{i+1}}$ has already been proven by lemma 4.5.3. Recall that the sequences $(s_i)_i$ and $(r_i)_i$ geometrically decays to zero, which directly implies that domains D_i shrink geometrically fast to the domain D_∞. Lastly, we add $\Phi_0 = 0$ to the sequence $(\Phi_i)_i$.

Therefore, all prerequisites for the Whitney inverse approximation lemma 2.3.3 have been met, and we apply it to prove that there exists a unique limit function Φ_∞ on D_∞ such that $(\Phi_i)_i$ uniformly converges to a Whitney-smooth Φ_∞ on D_∞.

4.5.3 Construction of Φ

Before we continue with the construction of Φ, we first apply the Whitney extension theorem 2.3.4 to the Whitney-smooth limit function Φ_∞ on D_∞.

This results in the extended Whitney smooth limit function

$$\tilde{\Phi}_\infty : \mathrm{cl}\,[U_\kappa(T^n)] \times \{0\} \times \Sigma \to \mathrm{cl}\,[U_\kappa(T^n)] \times \{0\} \times \Sigma. \tag{4.80}$$

Lastly, recall that each approximate conjugation Ψ_i is a linear function in the y-variable by construction we have used[10].

Since every composition of linear functions is linear itself, we have that each Φ_i, and in extension Φ_∞, is linear in the y-variable. Lastly, since the Whitney extension theorem preserves this linearity as it only acts on the ω parameter, we have that the extended limit function $\bar{\Phi}_\infty$ is linear in the y-variable aswell.

This linearity can be exploited to extend the extended Whitney smooth limit function $\tilde{\Phi}_\infty$, which is only defined for $\mathrm{cl}\,[U_\kappa(T^n)] \times \{0\} \times \Sigma$, to the final diffeomorphism

$$\Phi : \mathrm{cl}\,[U_\kappa(T^n)] \times \mathbb{C}^n \times \Sigma \to \mathrm{cl}\,[U_\kappa(T^n)] \times \mathbb{C}^n \times \Sigma. \tag{4.81}$$

The diffeomorphism Φ satisfies the following:

- Φ is a Whitney-smooth diffeomorphism, as both the parameter shift and Φ_ω are Whitney-smooth by construction.
- Φ is a symplectomorphism for fixed ω, since Φ_ω is a symplectomorphism by definition.
- Φ is C^∞-close to the identity, since Φ_ω preserves the estimates made in lemma 4.5.1 by the uniform convergence of $(\Phi_i)_i$ to Φ_∞.
- On $U_\kappa(T^n) \times \mathbb{C}^n \times \Sigma_1' \cap \Phi^{-1}(U_\kappa(T^n) \times 0 \times U_1(\Sigma))$, the Hamiltonian $H \circ \Phi = N_\infty$ is integrable and in the form (4.11), because Φ preserves the estimates given in the KAM step lemma 4.4.1. This is implied by the uniform convergence of $(\Phi_i)_i$ to Φ_∞.

Therefore, Φ satisfies all requirements set in theorem 4.2.2.

To fully prove the parameterized KAM theorem, only two steps remain: firstly, we must prove the existence of sequences that satisfy all assumptions, and secondly, we must reintroduce the γ parameter.

[10] This is a consequence of solving the homological equation (4.40) in each KAM step with the *linearized* perturbation P_{lin}.

4.5.4 Existence of the sequences

Throughout this proof, we have assumed many properties of the sequences $(s_i)_i$ and $(\delta_i)_i$. It is directly obvious that there exists $(s_i)_i$ and $(\delta_i)_i$ as defined in (4.13), which satisfy each of these assumptions.

Therefore, we verify that the sequences $(s_i)_i$ and $(\delta_i)_i$ can satisfy every assumption made in the proof of the KAM step lemma and the estimates on Φ_i.

Lemma 4.5.4. *There exist s_0 and δ_0 such that the sequences $(s_i)_i$ and $(\delta_i)_i$ as defined in* (4.13) *satisfy the basic conditions* (4.20) *and* (4.39)*, the KAM lemma conditions* (4.46)*,* (4.47)*,* (4.55)*,* (4.64) *and* (4.72)*, and lastly the convergence conditions* (4.76) *and* (4.78)*.*

For a proof of this lemma consult the appendix (lemma A.4.1). This lemma proves the parameterized KAM theorem 4.2.2 for the specific case $\gamma = 1$.

In modern KAM literature, it is common to omit the explicit verification of the sequences, since the assumptions are very similar across many KAM proofs. As KAM proofs are already lengthy, this step is often compressed to a few sentences to reduce the overall length of the proof.

4.5.5 Re-scaling to reintroduce γ

The very last step of this KAM proof is the re-introduction of γ by means of a suitable scaling. Consider the transformation

$$(x, y, \omega) \mapsto (X, Y, \Omega) = (x, y, \gamma^{-1}\omega), \tag{4.82}$$

and the time scaling $t = \gamma t$. Under this rescaling the normal form of N becomes

$$\gamma N = \langle \gamma^{-1}\omega, Y \rangle = \langle \Omega, Y \rangle. \tag{4.83}$$

This concludes the proof of theorem 4.2.2.

To apply the Whitney inverse approximation lemma (cf. lemma 2.3.3), to a sequence of functions $(f_i)_i$, we require a geometrically shrinking sequence of domains W_i on which $|f_{i+1} - f_i|_{W_{j+1}}$ decays super-exponentially.

Recall that the sequences $(s_i)_i$ and $(r_i)_i$ geometrically decays to zero. The sequence $(\varepsilon_i)_i$ super-exponentially decays, and can easily be dominated by a geometrically decaying sequence. Thus we can claim that the domains D_i shrink geometrically fast to the domain $U_\kappa(T^n) \times \{0\} \times \Sigma_1'$. This sequence of domains satisfies the definition 2.37. The super-exponential decay of $|\Phi_{i+1} - \Phi_i|_{D_{i+1}}$ has been proven by lemma 4.5.3. Lastly, we add $\Phi_0 = 0$ to the sequence $(\Phi_i)_i$.

All prerequisites for the Whitney inverse approximation lemma 2.3.3 have been met and we apply it to prove that there exists

a unique function Φ_∞ on $D_\infty = U_\kappa(T^n) \times \{0\} \times \Sigma_1'$, which satisfies the following;

- The sequence $(\Phi_i)_i$ uniformly converges to Φ_∞ on D_∞.
- Φ_∞ is smooth in the Whitney sense, or equivalently Φ_∞ is a Whitney-regular diffeomorphism, on D_∞.
- $H \circ \Phi_\infty = N_\infty$, where N_∞ is integrable and in the form (4.11).
- Φ_∞ is a symplectomorphism for fixed ω.
- Φ_∞ is C^∞-close to the identity.

The last three point follow from the uniform convergence of $(\Phi_i)_i$ and the properties of each Φ. (See section 4.3.2)

4.6 Classical KAM theorem

The classical KAM theorem as stated in the introduction of this section is a corollary of the parameterized KAM theorem 4.2.2. This is easily demonstrated by mapping the action variables y to the frequency parameter ω.

Frequency map

We re-parameterize ω by the internal parameters y, resulting in

$$\omega : R^n \to R^n : y \mapsto \omega(y), \tag{4.84}$$

where ω is a real analytic map. Our new parameter space is the small neighborhood $\mathbb{Y} := \mathcal{O}_\sigma$ of the origin, which is compact. The new frequency space is given by $\Sigma := \omega(\mathbb{Y})$. Note that the parameter space $\mathbb{Y}$ is part of the phase space $T^n \times \mathbb{Y}$, as we only used internal parameters in this case.

Non-degeneracy condition

With a non-trivial frequency map, the standard non-degeneracy condition does not automatically hold. Thus, we must reintroduce some non-degeneracy condition on ω.

The frequency map must be submersive to fully pull the whole geometry of the frequency space Σ back into the parameter space, and thus into the phase space $T^n \times \mathcal{O}_\sigma$. This implies that the preimage of the subset of Diophantine frequencies, $\omega^{-1}(\Sigma'_\gamma)$, has a non-trivial measure within the parameter space $\mathbb{Y}$. And, thus the set of invariant Diophantine tori in $T^n \times \mathbb{Y}$, given by $T^n \times \omega^{-1}(\Sigma'_\gamma)$, has a non-trivial measure in the phase space. Moreover, by lemma 2.2.5 we actually have that the Diophantine tori fill the phase space up to a set of measure $O(\gamma)$.

Thus a sufficient non-degeneracy condition is requiring ω to be submersive on $\mathbb{Y}$., i.e. the standard or BHT non-degeneracy condition. However, classical KAM theory is usually stated with the traditional Kolmogorov non-degeneracy condition, which is given by

$$\det D\omega(0) \neq 0. \tag{4.85}$$

As was mentioned before in section 3.5, this non-degeneracy condition implies that ω is a local diffeomorphism and thus satisfies the standard non-degeneracy condition for a sufficiently small neighbourhood $\mathbb{Y}$ of zero.

Proof of classical KAM theorem

What remains to be proven is whether the Diophantine tori survive perturbation, which is done by applying theorem 4.2.2.

We are given the perturbed Hamiltonian $H = N + P$, where N is integrable, the frequency map ω is as in (4.84) and that ω satisfies the Kolmogorov's non-degeneracy condition (4.85). The Hamiltonian N, by definition of integrability and the frequency map ω, has the form

$$N = \langle \omega(y), y \rangle. \tag{4.86}$$

This integrable Hamiltonian cannot be directly used in theorem 4.2.2 and we must adapt it. Firstly, we expand the frequency map about the origin in the y variable to obtain $\omega(y) = \omega + \omega_\epsilon(y)$, where $\omega_\epsilon(y)$ is a small remainder[11]. Secondly, we add the remainder $\langle \omega_\epsilon(y), y \rangle$ to the perturbation P, which brings N to the form required by theorem 4.2.2.

Next, we define $\mathbb{Y}$ small enough such that, firstly, the norm of $\langle \omega_\epsilon(y), y \rangle$ and $O(y^2)$ is sufficiently small, and secondly, that the domain D given by

$$D = U_\kappa(T^n) \times \mathbb{Y} \times \Sigma, \tag{4.87}$$

where $\Sigma = \omega(\mathbb{Y})$, when extended to the complex domain satisfies the requirements of theorem 4.2.2.

We can now apply theorem 4.0.2 to state that for a given $H = N + P$, there exists a constant $\tilde{\delta} > 0$ such that if $|P|_D < \tilde{\delta}$, then the Diophantine quasi-periodic invariant tori of the unperturbed system in D with frequencies in Σ'_γ survive the perturbation, being only slightly distorted.

Lastly, to obtain the constant δ as stated in theorem 4.0.1, we subtract the supremum norm of $\langle \omega_\epsilon(y), y \rangle$ on the small domain $\mathbb{Y}$. This is why we required the norm of $\langle \omega_\epsilon(y), y \rangle$ to be sufficiently small for all $y \in \mathbb{Y}$.

This concludes the proof of the classical KAM theorem 4.0.1.

[11] Since we are operating on a very small neighbourhood $\mathbb{Y}$ for the y variable.

Proofs of used lemmas

The KAM proof employs several lemmas for estimating the KAM step, the convergence of Φ, the existence of sequences. Proving these lemmas within the KAM proof itself would have disrupted the flow of the proof. Therefore, these lemmas are proved separately and presented here. This appendix also contains the formal solution to the linearized conjugation relations.

A.1 Formal solution to linearized conjugation relation.

We begin the solution by considering the Poisson bracket. We have

$$\{N, F\} = -\omega \frac{\partial}{\partial x} F, \tag{A.1}$$

as $\frac{\partial}{\partial x} N = 0$ and $\frac{\partial}{\partial y} N = \omega$, given that N is in normal form (4.11). Thus, setting $\bar{N} = N_+ - N$, the linearized conjugating relation (4.40) is equivalent to

$$P = \omega \frac{\partial}{\partial y} F + \bar{N}. \tag{A.2}$$

We consider the Taylor series expansion in the y variable about 0, and the Fourier series in the x variable for both P and F.

$$P(x, y;\, \omega) = \sum_{|j|\geq 0} \sum_{k\in\mathbb{Z}^n} P_k^j(\omega) e^{i\langle k,\, x\rangle} y^j \quad \text{and} \quad F(x, y;\, \omega) = \sum_{|j|\geq 0} \sum_{k\in\mathbb{Z}^n} F_k^j(\omega) e^{i\langle k,\, x\rangle} y^j. \tag{A.3}$$

Using these series, we have

$$\omega \frac{\partial}{\partial y} F = \sum_{|j|\geq 0} \sum_{k\in\mathbb{Z}^n} i\langle k,\, \omega\rangle F_k^j(\omega) e^{i\langle k,\, x\rangle} y^j. \tag{A.4}$$

We solve (A.2), by setting

$$P_k^j(\omega) = i\langle k, \omega\rangle F_k^j(\omega) \quad \text{for all} \quad j \text{ and } k \neq 0. \tag{A.5}$$

For $k = 0$, the term $i\langle k, \omega\rangle$ vanishes, and we have $\sum_{|j|\geq 0} P_0^j(\omega) e^{i\langle k, x\rangle} = \bar{N}$.

Note that $\bar{N}$ is not uniquely determined by this equation and can be chosen in various ways. However, for convenience, we choose the simplest one, which sets $\bar{N}$ to be equal to the average of P over T^n.

$$\bar{N} = [P(\cdot, y, \omega)] = \frac{1}{\mathrm{Vol}(T^n)} \int_{T^n} P(x, y, \omega) dx = e(\omega) + \langle v(\omega),\, y\rangle + \Xi, \tag{A.6}$$

where the higher order term in the y-variable are collected in Ξ, which is on the order of $O(y^2)$.This results in a new integrable part N_+

$$N_+ = N - \bar{N} = N + [P] = e(\omega) + \langle \omega + v(\omega),\, y\rangle + \Xi. \tag{A.7}$$

Therefore, to obtain the Hamiltonian F we must solve

$$F_k^j(\omega) = \frac{P_k^j(\omega)}{i\langle k,\, \omega\rangle} \quad \text{and} \quad N_+ = N + [P(\cdot, y, \omega)], \tag{A.8}$$

which requires us to manage the small denominators $i\langle k,\, \omega\rangle$.

Formally solving the new perturbation term

Let us assume we have constructed a sufficient F, and have used this Hamiltonian to construct the time-one map ϕ_1^F. This map is a Hamiltonian diffeomorphism by definition, and thus we can use (3.30) to calculate $H \circ \phi_1^F$

Applying this transformation to H and subtracting the new integrable part N_+ results in the new perturbation term:

$$P_+ = H \circ \phi_1^F - N_+ = (N+P) \circ \psi_1^F - N_+ = N \circ \psi_1^F + P \circ \psi_1^F - N_+ \tag{A.9}$$

$$= N + \{N, F\} + \int_0^1 (1-t) \cdot \{\{N, F\}, F\} \circ \psi_t^F \, dt + P + \int_0^1 \{P, F\} \circ \psi_t^F \, dt - N_+. \tag{A.10}$$

where we used the second and first order approximation with remainder for $N \circ \psi_1^F$ and $P \circ \psi_1^F$ respectively. We can simplify this further by reordering and recalling that $N + \{N, F\} - N_+ + P = 0$, resulting in

$$N + \{N, F\} - N_+ + P + \int_0^1 (1-t) \cdot \{\{N, F\}, F\} \circ \psi_t^F \, dt + \int_0^1 \{P, F\} \circ \psi_t^F \, dt \tag{A.11}$$

$$= \int_0^1 \{(1-t) \cdot \{N, F\} + P, F\} \circ \psi_t^F \, dt \tag{A.12}$$

$$= \int_0^1 \{(1-t) \cdot \bar{N} + tP, F\} \circ \psi_t^F \, dt. \tag{A.13}$$

where we used that $\{(1-t) \cdot N, F\} + P = (1-t) \cdot \bar{N} + tP$.

A.2 KAM step lemmas

In this section, we provide the proofs for all the lemmas used in the KAM step.

Truncation $P_{lin,d}$

Lemma A.2.1 (Truncation estimates)**.** *Under the assumptions of lemma 4.4.2, the truncation $P_{lin,d}$, as given in* (4.44)*, satisfies*

$$|P_{lin,d}|_{D^1} \lessdot \delta, \tag{A.14}$$

$$|\hat{P}|_{D_*^1} = |P - P_{lin,d}|_{D_*^1} \lessdot \varepsilon^\chi \delta. \tag{A.15}$$

Proof. We begin by proving the estimate $|P - P_{\mathrm{lin},d}|_{D_*^1} \lessdot \varepsilon^\chi \delta$. Firstly, observe that $P - P_{\mathrm{lin},d}$ can be split into two terms.

$$P - P_{\mathrm{lin},d} = \sum_{|j| \leq 1} \sum_{|k| > d} P_j^k(\omega) e^{2\pi \langle k, x \rangle} y^j + \sum_{|j|>1} P_j(x, \omega) y^j. \tag{A.16}$$

The first term is estimated via the Paley-Wiener estimate $\left|P_j^k(\omega) y^j\right| \leq |P|_D e^{-(\kappa+s)|k|}$ for all $k > d$.

$$\left| \sum_{|j| \leq 1} \sum_{|k|>d} P_j^k(\omega) y^j e^{i\langle k, x \rangle} \right|_{D_*^1} \lessdot \sum_{|k|>d} |P|_D e^{-(\kappa+s)|k|} e^{|\mathrm{Im}(x)| \cdot |k|} \leq \delta \sum_{|k|>d} e^{-\frac{5}{8} s|k|}, \tag{A.17}$$

where we used $|P|_{D_*^1} \leq |P|_D \leq \delta$ and $|\mathrm{Im}(x)| \leq \kappa + \frac{3}{8}s$ on D_*^1. We proceed by estimating the sum.

$$\delta \sum_{|k|>d} e^{-\frac{5}{8} s|k|} \lessdot d^n e^{-\frac{5}{8} sd} \delta, \tag{A.18}$$

where we used lemma A.5.3 and assumption (4.47). Lastly. under the assumption that $d^n e^{-\frac{5}{8} ds} < \varepsilon$ (cf. assumption (4.46)), we have that

$$d^n e^{-\frac{5}{8} sd} < \varepsilon < \varepsilon^\chi, \tag{A.19}$$

which finishes the estimate of the first term.

The estimate of the second term is made with use of the remainder formula[1] of the Taylor expansion about $y = 0$ on the domain D^1_*. The remainder consists of terms in the form

$$R_\alpha(x, y, \omega) = \frac{|\alpha|}{\alpha!} y^\alpha \int_0^1 (1-t)^{|\alpha|-1} \partial y^\alpha P(x, ty, \omega)\, dt. \tag{A.20}$$

Each R_α has the property that $|R_\alpha|_{D^1_*} \leq \frac{1}{\alpha!} |\partial y^\alpha P|_{D^1_*} \cdot |y^\alpha|_{D^1_*}$. We see that

$$P - P_{\text{lin}} = \sum_{|j|>1} P_j(x, \omega) y^j = \sum_{|j|=2} R_j(x, y, \omega), \tag{A.21}$$

and consequently that

$$\left| \sum_{|j|>1} P_j(x, \omega) \right|_{D^1_*} = \left| \sum_{|j|=2} R_j(x, y, \omega) \right|_{D^1_*} \leq \sum_{|j|=2} \frac{1}{j!} \left| \partial y^j P \right|_{D^1_*} \cdot |y^\alpha|_{D^1_*} \tag{A.22}$$

Recall that y is bounded by $8\beta\varepsilon = 8\varepsilon^{\chi/2}\varepsilon$ on D^1_* and bounded by ε on D. We apply Cauchy's estimate and the maximum bound of a polynomial to obtain

$$|\partial y^\alpha P|_{D^1_*} \leq \alpha!(\varepsilon - \beta\varepsilon)^{-|\alpha|}\delta \quad \text{and} \quad |y^\alpha|_{D^1_*} \leq (\beta\varepsilon)^{|\alpha|}, \tag{A.23}$$

which if substituted in (A.22) result in

$$\sum_{|j|=2} \frac{1}{j!} \left| \partial y^j P \right|_{D^1} \cdot \left| y^j \right|_{D^1} \leq \delta \sum_{|j|=2} (\varepsilon - \beta\varepsilon)^{-|j|} (\beta\varepsilon)^{|j|} \lessdot (\varepsilon - \beta\varepsilon)^{-2} (\beta\varepsilon)^2 \delta = \frac{\beta^2}{1 + \beta^2 - 2\beta} \delta. \tag{A.24}$$

Lastly, since $0 < \beta < \frac{1}{2}$, we have

$$\frac{\beta^2}{1 + \beta^2 - 2\beta} \delta = \frac{1}{(1-\beta)^2} \beta^2 \delta \leq 4\beta^2 \delta \lessdot \varepsilon^\chi \delta, \tag{A.25}$$

which, together with (A.18) and (A.19), proves the estimate (A.15).
To prove the estimate (A.14), we can repeat the above argument with the domain D^1 to obtain

$$|P - P_{\text{lin},d}|_{D^1} \lessdot \delta, \tag{A.26}$$

which by $|P_{\text{lin},d}|_{D^1} \leq |P - P_{\text{lin},d}|_{D^1} + |P|_{D^1}$ directly leads to (A.14). □

[1] Consult the book [98] for a proof of the multivariate Taylor expansion.

Diophantine frequencies

Lemma A.2.2 (Small denominator estimate)**.** *For all $\omega \in U_r(\Sigma_1')$ and $k \in \mathbb{Z}^n$ with $0 < |k| \leq d$ one has*

$$|\langle \omega, k \rangle| \geq \tfrac{1}{2}|k|^{-\tau}. \tag{A.27}$$

Proof. Recall that $r = \frac{1}{2}s^{2\tau+2}$ and $d = [s^{-2}] \leq s^{-2}$, and therefore that

$$0 < rd \leq \tfrac{1}{2}s^{2\tau+2}s^{-2} = \tfrac{1}{2}s^{2\tau} \leq \tfrac{1}{2}d^{-\tau} \leq \tfrac{1}{2}|k|^{-\tau}. \tag{A.28}$$

For every $\omega \in U_r(\Sigma_1')$, there exists an $\omega^* \in \Sigma_1'$ such that $|\omega - \omega^*| < r$, thus we can state

$$|\langle \omega, k \rangle| \geq |\langle \omega^*, k \rangle| - |\omega - \omega^*| \cdot |k| \geq |k|^{-\tau} - rd \geq |k|^{-\tau} - \tfrac{1}{2}|k|^{-\tau} \geq \tfrac{1}{2}|k|^{-\tau}. \tag{A.29}$$

□

Hamiltonian function F

Lemma A.2.3 (Estimate on F)**.** *The Hamiltonian function F satisfies*

$$\begin{array}{cccccc} s^{2\tau+1}|F|_{D^2}, & s^{2\tau+2}\,|\partial_x F|_{D^3}\,, & \varepsilon s^{2\tau+1}\,|\partial_y F|_{D^3} & \lessdot & \delta, \\ s^{2\tau+3}\,|\partial_{xx} F|_{D^4}\,, & \varepsilon s^{2\tau+2}\,|\partial_{xy} F|_{D^4}\,, & \varepsilon s^{2\tau+2}\,|\partial_{yx} F|_{D^4} & \lessdot & \delta. \\ & r s^{2\tau+2}\,|\partial_{x\omega} F|_{D^4}\,, & r\varepsilon s^{2\tau+2}\,|\partial_{y\omega} F|_{D^4} & \lessdot & \delta. \end{array} \tag{A.30}$$

and lastly $\partial_{yy} F = 0$.

Proof. Recall that F is determined by

$$\sum_{0<|k|\leq d} \sum_{|j|\leq 1} F_k^j(\omega) e^{i\langle k,\, x\rangle} \quad \text{with} \quad F_k^j(\omega) = \frac{P_k^j(\omega)}{i\langle k,\, \omega\rangle}, \tag{A.31}$$

and, therefore, that the norm of F on D^2 is estimated by

$$\begin{aligned} |F|_{D^2} &= \left| \sum_{0<|k|\leq d} \sum_{|j|\leq 1} \frac{P_k^j(\omega) y^j}{i\langle k,\, \omega\rangle} e^{i\langle k,\, x\rangle} \right|_{D^2} \leq \sum_{0<|k|\leq d} \sum_{|j|\leq 1} \frac{\left|P_k^j(\omega) y^j\right|_{D^2}}{|i\langle k,\, \omega\rangle|} \left| e^{i\langle k,\, x\rangle} \right|_{D^2} \\ &\lessdot \delta \sum_{0<|k|\leq d} |k|^{\tau} e^{-(\kappa+s)|k|} e^{|\mathrm{Im}(x)|\cdot|k|} = \delta \sum_{0<|k|\leq d} |k|^{\tau} e^{-\frac{2}{8}s|k|}, \end{aligned}$$

where we used lemma A.2.2, the following Paley-Wiener estimate $\left|P_j^k(\omega) y^j\right|_{D^2} \leq e^{-(\kappa+s)|k|}|P|_D$, and lastly that $|\mathrm{Im}(x)| \leq \kappa + \frac{6}{8}s$ on D^2. We apply corollary A.5.2 to conclude that

$$|F|_{D^2} \lessdot \delta \sum_{0<|k|\leq d} |k|^{\tau} e^{-\frac{2}{8}s|k|} \lessdot s^{-2\tau-1}\delta. \tag{A.32}$$

The rest of the estimates directly follow from applying Cauchy's estimate multiple times and the estimate of $|F|_{D^2}$. Finally, we see that F by (A.31) is at most linear in the y variable, which automatically results in $\partial_{yy} F = 0$. □

As a comment to this proof, one should note since $P_{\mathrm{lin},d}$ was linearized in y, so is F and the vector field that is induced by it.

Transformation ψ_ω

Lemma A.2.4. *Under the assumptions of lemma 4.4.5, we have that the transformation*

$$\psi_\omega : D_*^2 \to D_*^1 : (x, y, \omega) \mapsto (\varphi_V^1(x, y), \omega) \tag{A.33}$$

is well-defined and satisfies

$$s^{2\tau+2}|\psi_\omega - id|_{D_*^2} <\cdot \varepsilon^{-1}\delta, \tag{A.34}$$

$$rs^{2\tau+2}|D\psi_\omega - Id|_{D_*^2} <\cdot \varepsilon^{-1}\delta, \tag{A.35}$$

where id is the identity map and Id is the identity matrix.

Proof. To begin, we define the domains $\tilde{D}_*^\iota = U_{\kappa+s_\iota^*}(T^n) \times \mathcal{O}_{\xi_\iota^*}$. To prove that ψ_ω is well-defined, we must show that it maps the intermediate domain D_*^2 to $\tilde{D}_*^1$. This is equivalent to proving that the time-one flow map of the vector field V_ω maps $\tilde{D}_*^2$ to $\tilde{D}_*^1$. Recall that V_ω is given by

$$V_\omega = \partial_y F \frac{\partial}{\partial x} - \partial_x F \frac{\partial}{\partial y}, \tag{A.36}$$

and that its flow map φ_V^t satisfies

$$\varphi_V^t = \mathrm{id} + \int_0^t V_\omega \circ \varphi_V^{\bar{t}} d\bar{t}. \tag{A.37}$$

Therefore, assuming that t is small enough that φ_V^t maps $\widetilde{D}_*^2$ to $\widetilde{D}_*^1$, the maximum displacement of the flow map φ_V^t is given by

$$|\varphi_V^t - \mathrm{id}|_{\tilde{D}_*^2} = \left| \int_0^t V_\omega \circ \varphi_V^{\bar{t}} d\bar{t} \right|_{\tilde{D}_*^2} \leq \int_0^t \left| V_\omega \circ \varphi_V^{\bar{t}} \right|_{\tilde{D}_*^2} d\bar{t} \leq \int_0^t |V_\omega|_{\tilde{D}_*^1} \, d\bar{t} = |V_\omega|_{\tilde{D}_*^1} |t|. \tag{A.38}$$

We calculate the x and y displacements of the time-one flow map φ_V^t separately. We estimate $|(V_\omega)_x|_{\tilde{D}_*^1}$ and $|(V_\omega)_y|_{\tilde{D}_*^1}$ for all relevant ω by considering them in the intermediate domain D^3. We see that

$$|(\varphi_V^1 - \mathrm{id})_y|_{\tilde{D}_2^*} \leq |\partial_x F|_{D^3} < cs^{-2\tau-2}\delta < \varepsilon_+ \quad \text{and} \quad |(\varphi_V^1 - \mathrm{id})_x|_{\tilde{D}_2^*} \leq |\partial_y F|_{D^3} < c\varepsilon^{-1}s^{-2\tau-1}\delta < s, \tag{A.39}$$

where we applied lemma A.2.3 with the constant made explicit and used the assumption that $c\delta < \varepsilon_+ s^{2\tau+2} < \varepsilon s^{2\tau+2}$ (cf. assumption 4.64). The inequalities (A.39) directly imply that φ_V^1 maps $\tilde{D}_*^2$ into $\tilde{D}_*^1$. Lastly, the total displacement can be estimated by

$$|\varphi_V^1 - \mathrm{id}|_{\tilde{D}_*^2} \leq |V_\omega|_{\tilde{D}_*^1} \leq |V_\omega|_{D^3} = \max\left(|\partial_x F|_{D^3}, |\partial_y F|_{D^3}\right) <\cdot \max\left(s^{-2\tau-2}\delta, \varepsilon^{-1}s^{-2\tau-1}\delta\right) = \varepsilon^{-1}s^{-2\tau-1}\delta.$$

Recall that $\psi_\omega : (x, y, \omega) \mapsto \left(\varphi_V^1(x, y), \omega\right)$. From the previous arguments, we see that ψ_ω indeed maps D_*^2 to D_*^1 and that $s^{2\tau+1}|\psi_\omega - \mathrm{id}|_{D_*^2} <\cdot \varepsilon^{-1}\delta$.

We continue with the estimate on the total derivative, which analogously to the estimate of the displacement is estimated for all relevant ω on the much larger intermediate domain D^4. This total derivative must also consider the ω dependence of ψ_ω.

$$|D\psi_\omega - \mathrm{Id}|_{D_*^2} \leq |D\psi_\omega - \mathrm{Id}|_{D^4} = \left| \begin{pmatrix} \partial_{xy}F & \partial_{xx}F & \partial_{x\omega}F \\ \partial_{yy}F & \partial_{yx}F & \partial_{y\omega}F \\ 0 & 0 & 0 \end{pmatrix} \right|_{D^4} <\cdot \max \begin{pmatrix} |\partial_{xy}F|_{D^4}, & |\partial_{xx}F|_{D^4}, & |\partial_{x\omega}F|_{D^4} \\ |\partial_{yy}F|_{D^4}, & |\partial_{yx}F|_{D^4}, & |\partial_{y\omega}F|_{D^4} \end{pmatrix} \tag{A.40}$$

We apply lemma A.2.3 to obtain

$$|D\psi_\omega - \mathrm{Id}|_{D_*^2} <\cdot \max \begin{pmatrix} \varepsilon^{-1}s^{-2\tau-2}\delta, & s^{-2\tau-3}\delta, & r^{-1}s^{-2\tau-2}\delta \\ 0, & \varepsilon^{-1}s^{-2\tau-2}\delta, & \varepsilon^{-1}r^{-1}s^{-2\tau-2}\delta \end{pmatrix} = \varepsilon^{-1}r^{-1}s^{-2\tau-2}\delta. \tag{A.41}$$

Therefore, we conclude that $rs^{2\tau+2}|D\psi_\omega - \mathrm{Id}|_{D_*^2} <\cdot \varepsilon^{-1}\delta$.

□

Parameter shift ϕ

Lemma A.2.5 (Parameter shift ϕ lemma)**.** *Under the assumptions of lemma 4.4.6, we have that the parameter shift*

$$\phi : D^3_* \to D^2_* : (x, y, \omega_+) \mapsto (x, y, \omega), \tag{A.42}$$

is well-defined and satisfies

$$|\phi - id|_{D^3_*} \lessdot \varepsilon^{-1}\delta, \tag{A.43}$$

$$r|D\phi - Id|_{D^3_*} \lessdot \varepsilon^{-1}\delta, \tag{A.44}$$

where id is the identity map, and Id is the identity matrix.

Proof. We begin by proving that

$$|\omega - \omega_+| = |P^1_0(\omega)| \leq \frac{1}{8}r \tag{A.45}$$

for all $\omega_+ \in U_{r^*_3}(\Sigma'_1)$, as this implies that ϕ maps every $\omega_+ \in U_{r^*_3}(\Sigma'_1)$ to an $\omega \in U_{r^*_2}(\Sigma'_1)$. And, hence, that ϕ maps D^3_* to D^2_*.

For the moment, we assume that ϕ maps ω_+ to an $\omega \in D^2_*$, and thus that

$$|P^1_0(\omega)|_{D^2_*} \leq |P^1_0(\omega)|_{D^1} = \max_j \left|P^j_0(\omega)\right|_{D^1} \lessdot |\partial_y P(0,\omega)|_{D^1} \lessdot \varepsilon^{-1}|P|_d = \varepsilon^{-1}\delta \leq \frac{1}{8}r. \tag{A.46}$$

where we used the assumption $8\delta \leq \varepsilon r$ (cf. assumption (4.64)). The inequality (A.46) directly implies that ϕ maps D^3_* into D^2_* and justifies the assumption that $\omega \in D^2_*$. It also implies that $|\phi - \mathrm{id}|_{D^3_*} \lessdot \varepsilon^{-1}\delta$ since $|\phi - \mathrm{id}|_{D^3_*} \leq |P^0_1(\omega)|_{D^2_*}$. We continue with the total derivative.

$$|D\phi - \mathrm{Id}|_{D^3_*} = \left|\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \partial_\omega\phi - \mathrm{Id} \end{pmatrix}\right|_{D^3_*} \lessdot |\partial_\omega\phi - \mathrm{Id}|_{D^3_*} \leq r^{-1}|\phi - \mathrm{id}|_{D^1} \lessdot r^{-1}\varepsilon^{-1}\delta, \tag{A.47}$$

where we applied Cauchy's estimate.

□

Complete transformation Ψ

Lemma A.2.6 (Estimates on Ψ)**.** *Under the assumptions of lemma 4.4.5 and 4.4.6, we have that the coordinate transformation*

$$\Psi : D^3_* \to D^1_* : \Psi := \psi_\omega \circ \phi, \tag{A.48}$$

is well-defined and satisfies

$$s^{2\tau+2}|\Psi - id\,|_{D^3_*} \lessdot \varepsilon^{-1}\delta, \tag{A.49}$$

$$rs^{2\tau+2}|D\Psi - Id\,|_{D^2_*} \lessdot \varepsilon^{-1}\delta, \tag{A.50}$$

where id is the identity map, and Id is the identity matrix.

Proof. This proof is relatively simple since Ψ is just the composition of the maps ψ_ω and ϕ. Firstly, by lemmas A.2.4 and A.2.5, we know that ψ_ω and ϕ are well-defined, and hence, so is Φ. The first estimate is proven by

$$|\Psi - \mathrm{id}\,|_{D^3_*} = |\psi_\omega \circ \phi - \mathrm{id}\,|_{D^3_*} \leq |\psi_\omega - \mathrm{id}\,|_{D^3_*} \leq s^{-2\tau-2}\varepsilon^{-1}\delta, \tag{A.51}$$

where we used lemma A.2.4. The second estimate is a consequence of the chain rule.

$$\begin{aligned}
|D\psi_\omega - \mathrm{Id}\,|_{D^3_*} &= |D(\psi_\omega \circ \phi) - \mathrm{Id}\,|_{D^3_*} && \text{(A.52)}\\
&= |D_\phi\psi_\omega \circ D\phi - \mathrm{Id}\,|_{D^3_*} && \text{(A.53)}\\
&= |D_\phi\psi_\omega \circ (D\phi - \mathrm{Id}) + D_\phi\psi_\omega - \mathrm{Id}\,|_{D^3_*} && \text{(A.54)}\\
&\leq |D_\phi\psi_\omega \circ (D\phi - \mathrm{Id})|_{D^3_*} + |D_\phi\psi_\omega - \mathrm{Id}\,|_{D^3_*} && \text{(A.55)}\\
&\leq |D\psi_\omega|_{D^2_*} \cdot |D\phi - \mathrm{Id})|_{D^3_*} + |D\psi_\omega - \mathrm{Id}\,|_{D^2_*} && \text{(A.56)}\\
&\lessdot \max\left(r^{-1}\varepsilon^{-1}\delta, r^{-1}s^{-2\tau-2}\varepsilon^{-1}\delta\right) \leq r^{-1}s^{-2\tau-2}\varepsilon^{-1}\delta, && \text{(A.57)}
\end{aligned}$$

where we used that $D\psi_\omega$ is close to the identity matrix, which follows from the smallness of $|D\psi_\omega - \mathrm{Id}|_{D^2_*}$.

□

New perturbation P_+ estimates

Lemma A.2.7. *Under the assumptions of lemma 4.4.8, the new perturbation term P_+ satisfies*

$$|P_+|_{D_+} \leq \delta_+. \tag{A.58}$$

Proof. Recall that P_+ is given by

$$P_+ = (P - P_{\text{lin},,d}) \circ \psi_\omega \circ \phi + \left(\int_0^1 \{(1-t)\bar{N} + tP_{\text{lin},d}, F\} \circ \varphi_t^V \, dt \right) \circ \phi, \tag{A.59}$$

where $\bar{N} = N_+ - N = \langle P_1^0(\omega), y \rangle$. The first term can be easily estimated as follows:

$$|(P - P_{\text{lin},d}) \circ \psi_\omega \circ \phi|_{D_*^3} \leq |(P - P_{\text{lin},d}) \circ \psi_\omega|_{D_*^2} \leq |P - P_{\text{lin},d}|_{D_*^1} <\cdot \varepsilon^\chi \delta, \tag{A.60}$$

where we used lemma A.2.1.

The estimate of the integral is more involved. We must first consider the estimates on the norm of the Poisson brackets $\{P_{\text{lin},d}, F\}$ and $\{\bar{N}, F\}$ on D_*^1. To do so, we consider them on the much larger intermediate domain D^3.

$$\begin{aligned}
|\{P_{\text{lin},d}, F\}|_{D_*^1} &\leq |\{P_{\text{lin},d}, F\}|_{D^3} && \text{(A.61)}\\
&<\cdot \, |\partial_y P_{\text{lin},d}|_{D^3} \cdot |\partial_x F|_{D^3} - |\partial_x P_{\text{lin},d}|_{D^3} \cdot |\partial_y F|_{D^3} && \text{(A.62)}\\
&<\cdot \frac{\delta}{\varepsilon} \cdot \frac{\delta}{s^{2\tau+2}} + \frac{\delta}{s} \cdot \frac{\delta}{\varepsilon s^{2\tau+1}} <\cdot \frac{\delta^2}{\varepsilon s^{2\tau+2}}, && \text{(A.63)}
\end{aligned}$$

where we used the estimates of $|\partial_x F|_{D^3}$ and $|\partial_y F|_{D^3}$ from lemma A.2.3. In addition, we used the estimate $|P_{\text{lin},d}|_{D^1} <\cdot \delta$ from lemma A.2.1, along with Cauchy's estimate, to estimate $|\partial_x P_{\text{lin},d}|_{D^3}$ and $|\partial_y P_{\text{lin},d}|_{D^3}$.

We proceed with estimating $|\{\bar{N}, F\}|_{D_*^1}$.

$$\begin{aligned}
|\{\bar{N}, F\}|_{D_*^1} &\leq |\{\bar{N}, F\}|_{D^3} && \text{(A.64)}\\
&<\cdot \, |\partial_y \bar{N}|_{D^3} \cdot |\partial_x F|_{D^3} - |\partial_x \bar{N}|_{D^3} \cdot |\partial_y F|_{D^3} && \text{(A.65)}\\
&<\cdot \frac{\delta}{\varepsilon} \cdot \frac{\delta}{s^{2\tau+2}} + \frac{\delta}{s} \cdot \frac{\delta}{\varepsilon s^{2\tau+1}} <\cdot \frac{\delta^2}{\varepsilon s^{2\tau+2}}, && \text{(A.66)}
\end{aligned}$$

where we again used the estimates of $|\partial_x F|_{D^3}$ and $|\partial_y F|_{D^3}$ from lemma A.2.3. In addition, we used the estimate $|\partial_y \bar{N}|_{D^3} = |P_1^0(\omega)|_{D^3} <\cdot \varepsilon^{-1}\delta$ and the fact that $\partial_x \bar{N} = 0$.

With the estimates (A.63) and (A.66), we obtain

$$\begin{aligned}
\left| \int_0^1 \{(1-t)\bar{N} + tP_{\text{lin},d}, F\} \circ \varphi_V^t \, dt \circ \phi \right|_{D_*^3} &\leq \left| \int_0^1 \{(1-t)\bar{N} + tP_{\text{lin},d}, F\} \circ \varphi_V^t \, dt \right|_{D_*^2} && \text{(A.67)}\\
&\leq \int_0^1 \left| \{(1-t)\bar{N} + tP_{\text{lin},d}, F\} \right|_{D_*^1} dt && \text{(A.68)}\\
&\leq |\{\bar{N}, F\}|_{D_*^1} \int_0^1 (1-t)dt + |\{P_{\text{lin},d}, F\}|_{D_*^1} \int_0^1 t dt && \text{(A.69)}\\
&\leq |\{\bar{N}, F\}|_{D_*^1} + |\{P_{\text{lin},d}, F\}|_{D_*^1} && \text{(A.70)}\\
&\leq c\frac{\delta^2}{\varepsilon s^{2\tau+2}} \leq \varepsilon^\chi \delta = \delta_+, && \text{(A.71)}
\end{aligned}$$

where we made the constant from the estimates (A.63) and (A.66) explicit, and used the assumption that $c\delta \leq \varepsilon^{1+\chi} s^{2\tau+2}$ (cf. assumption (4.72)).

Finally, by combining the estimates (A.60) and (A.71), we obtain

$$|P_+|_{D_+} \leq |P_+|_{D_*^3} \leq \varepsilon^\chi \delta = \delta_+, \tag{A.72}$$

which proves the lemma. □

A.3 Convergence lemmas

In this section, we provide the proofs for all the lemmas used in the convergence part of the KAM proof.

Norm of Φ_i and $\Phi_{i+1} - \Phi_i$

Lemma A.3.1 (Estimates of $||\Phi_+||_{1,D_{i+1}}$)**.** *The transformation Φ_{i+1} satisfies*

$$||\Phi_{i+1}||_{1,D_{i+1}} \lesssim ||\Phi_i||_{1,D_i} \cdot \max\left(1 + |\Psi_i - id|_{D_{i+1}}, 1 + |D\Psi_i - Id|_{D_{i+1}}\right), \tag{A.73}$$

$$|\Phi_{i+1} - \Phi_i|_{D_{i+1}} \lesssim ||\Phi_i||_{1,D_i} \cdot |\Psi_i - id|_{D_{i+1}}, \tag{A.74}$$

for all $i \in \mathbb{N}_0$.

Proof. We use the plus notation and the intermediate domains D^ι and D^ι_* in this proof. Note that $\Phi_+ = (\Phi \circ \Psi - \Phi) + \Phi$. Therefore, we have

$$|\Phi_+|_{D_+} \leq |\Phi \circ \Psi - \Phi|_{D_+} + |\Phi|_{D_+} \leq |D\Phi|_{D^3_*} |\Psi - \mathrm{id}|_{D_+} + |\Phi|_{D_+} \leq |D\Phi|_D \left(1 + |\Psi - \mathrm{id}|_{D_+}\right), \tag{A.75}$$

where we have applied the mean value theorem to obtain the (point-wise) inequality

$$|\Phi \circ f(1) - \Phi \circ f(0)| \leq \max_{0<t<1} |D\Phi(f(t))| \cdot |f(1) - f(0)| \leq |D\Phi|_{D^3_*} \cdot |\Psi - \mathrm{id}|, \tag{A.76}$$

where $f(t) := (1-t)\,\mathrm{id} + t\Psi$. Note that $f(t) \in D^3_*$ as Ψ maps D_+ to D^3_*, and therefore that $\max_{0<t<1} |D\Phi(f(t))| \leq |D\Phi|_{D^3_*}$. Analogously, we obtain

$$|D\Phi_+|_{D_+} \leq |D(\Phi \circ \Psi) - D\Phi|_{D_+} + |D\Phi|_{D_+} \leq |D\Phi|_{D^4} \cdot |D\Psi - \mathrm{Id}|_{D_+} + |D\Phi|_{D_+} \leq |D\Phi|_D \cdot \left(1 + |D\Psi - \mathrm{Id}|_{D_+}\right). \tag{A.77}$$

By combining (A.3) and (A.77), we obtain the first estimate:

$$||\Phi_+||_{1,D_+} \leq ||\Phi||_{1,D} \cdot \max\left(|\Psi - \mathrm{id}|_{D_+}, |D\Psi - \mathrm{Id}|_{D_+}\right). \tag{A.78}$$

For the second estimate, we have

$$|\Phi_+ - \Phi|_{D_+} = |\Phi \circ \Psi - \Phi|_{D_+} \leq |D\Phi|_D \cdot |\Psi - \mathrm{id}|_{D_+}, \tag{A.79}$$

which follows directly from the calculations made in (A.3). □

Universal bound of $||\Phi_i||_{1,D_i}$.

Lemma A.3.2 (Universal bound of $||\Phi_i||_{1,D_i}$)**.** *Under the assumptions of lemma 4.5.2, the sequence $(||\Phi_i||_{1,D_i})_i$ is universally bounded, i.e. there exists a constant $C > 0$ such that*

$$||\Phi_i||_{1,D_i} \leq C \quad \textit{for all} \quad i \in \mathbb{N}_0. \tag{A.80}$$

Proof. By combining lemma A.3.1 and A.2.6, we obtain

$$||\Phi_{i+1}||_{1,D_{i+1}} \lesssim ||\Phi_i||_{1,D_i} \cdot \max\left(1 + |\Psi_i - \mathrm{id}|_{D_{i+1}}, 1 + |D\Psi_i - \mathrm{Id}|_{D_{i+1}}\right) \tag{A.81}$$

$$\leq ||\Phi_i||_{1,D_i} \left(1 + \max\left(s_i^{-2\tau-2}\varepsilon_i^{-1}\delta_i, r_i^{-1}s_i^{-2\tau-2}\varepsilon_i^{-1}\delta_i\right)\right) \tag{A.82}$$

$$\lesssim ||\Phi_i||_{1,D_i} \left(1 + \frac{\delta_i}{\varepsilon_i s_i^{4\tau+4}}\right). \tag{A.83}$$

Therefore, we have

$$||\Phi_{i+1}||_{1,D_{i+1}} \lesssim \prod_{j=0}^{i} \left(1 + \frac{\delta_j}{\varepsilon_j s_j^{4\tau+4}}\right), \tag{A.84}$$

which converges as $i \to \infty$ if the series $\sum_{j=0}^{\infty} \frac{\delta_j}{\varepsilon_j s_j^{4\tau+4}}$ converges. This series convergences if

$$\frac{\delta_{j+1}}{\varepsilon_{j+1} s_{j+1}^{4\tau+4}} < \frac{\delta_j}{\varepsilon_j s_j^{4\tau+4}} \tag{A.85}$$

holds for all $j \in \mathbb{N}_0$. This in turn holds if

$$\delta_0^{\chi/4} 4^{4\tau+4} < 1, \tag{A.86}$$

which is directly stated by the assumption (4.76).

□

Exponential decay of the coefficient functions

Lemma A.3.3 (Super-exponential decay of $|\Phi_{i+1} - \Phi_i|_{D_{i+1}}$)**.** *Under the assumptions of lemma 4.5.3, for* $0 < b < \frac{1}{2}$ *there exists a constant* $M > 0$ *such that*

$$|\Phi_{i+1} - \Phi_i|_{D_{i+1}} < M\delta_0^{b^i}. \tag{A.87}$$

i.e. the norm $|\Phi_{i+1} - \Phi_i|$ *on* D_{i+1} *converges super-exponentially fast to zero.*

Proof. By lemma A.3.1 , A.3.2 and A.2.6 we have

$$|\Phi_{i+1} - \Phi_i|_{D_{i+1}} \lesssim ||\Phi_i||_{1,D_i} \cdot |\Psi_i - \mathrm{id}|_{D_{i+1}} \lesssim s_i^{-2\tau-2}\varepsilon_i^{-1}\delta_i. \tag{A.88}$$

Therefore, $|\Phi_{i+1} - \Phi_i|_{D_{i+1}}$ decays super exponentially as soon as $s_i^{-2\tau-2}\varepsilon_i^{-1}\delta_i$ does. To ensure this, we choose a $0 < b < \frac{1}{2}$. We see that $s_i^{-2\tau-2}\varepsilon_i^{-1}\delta_i < \delta_i^b$ holds if for all $i \in \mathbb{N}_0$ we have

$$\delta_i^{1/2-b} < s_i^{2\tau+2}. \tag{A.89}$$

This condition is satisfied if

$$\delta_0^{1/2-b} < s_0^{2\tau+2} \quad \text{and} \quad \delta_0^{\chi/2(1/2-b)} < 4^{-2\tau-2}, \tag{A.90}$$

which is directly implied by the assumption (4.78). □

A.4 Sequence lemmas

In this section, we provide the proofs for all the lemmas used to prove the existence of suitable sequences $(s_i)_i$ and $(\delta_i)_i$.

Existence of the sequences

Lemma A.4.1. *There exist s_0 and δ_0 such that the sequences $(s_i)_i$ and $(\delta_i)_i$ as defined in* (4.13) *satisfy the basic conditions* (4.20) *and* (4.39)*, the KAM lemma conditions* (4.46)*,* (4.47)*,* (4.55)*,* (4.64) *and* (4.72)*, and lastly the convergence conditions* (4.76) *and* (4.78)*.*

Proof. The condition (4.20) is already stated on s_0 and δ_0, for convenience we repeat it here

$$0 < s_0 < \min\{\kappa, \tfrac{1}{2}\} \quad \text{and} \quad 0 < \delta_0 < \tfrac{1}{2}. \tag{A.91}$$

The condition (4.39) is satisfied if

$$\delta_0^{\chi/4} < \tfrac{1}{128}. \tag{A.92}$$

The condition (4.46) is the most involved, and thus we refer to lemma A.4.2, which states that (4.46) holds if

$$s_0 < \frac{1}{(c_1 + C) + c_2 \log(\delta_0^{-1})}, \tag{A.93}$$

where the constants c_1 and c_2 are explicitly given by (A.109), and C is a sufficiently large constant.
The condition (4.47) is satisfied if

$$s_0 \leq \frac{1}{\frac{8}{5}n + 1}. \tag{A.94}$$

The conditions (4.55) and (4.64) reduce to $c\delta \leq \varepsilon s^{2\tau+2}$ with a constant $c > 1$, which is satisfied if we have both

$$\delta_0^{1/2} \leq \tfrac{1}{c} s_0^{2\tau+2} \quad \text{and} \quad \delta_0^{\chi/2} \leq \tfrac{1}{c} \left(1/4\right)^{2\tau+2}. \tag{A.95}$$

The condition (4.72), given by $c\delta^{1-\chi} < s^{2\tau+2}$ with constant $c > 1$, is satisfied if we have both

$$\delta_0^{1-\chi} \leq \frac{1}{c} s_0^{2\tau+2} \quad \text{and} \quad \delta_0^{(1-\chi)\chi} \leq 4^{-(2\tau+2)}. \tag{A.96}$$

The conditions (4.76) and (4.78) are already stated in terms of s_0 and δ_0, resulting in

$$\delta_0^{\chi/4} < 4^{-4\tau-4}, \tag{A.97}$$

and for arbitrary $0 < b < \frac{1}{2}$

$$\delta_0^{1/2-b} < s_0^{2\tau+2} \quad \text{and} \quad \delta_0^{\chi/2(1/2-b)} < 4^{-2\tau-2}. \tag{A.98}$$

In the end we see that all conditions come in the forms

$$s_0 \leq c, \quad \delta_0 \leq c, \quad \delta_0^{\zeta} \leq c s_0 \quad \text{and} \quad s_0 < \frac{1}{c - \zeta \ln \delta_0}. \tag{A.99}$$

with constants $c > 0$ and exponents $\zeta > 0$ depending on the definitions of the sequences, n and τ. These conditions can all be satisfied at the same time, by choosing both δ_0 and s_0 small and in an interdependent way. This is due to

$$\lim_{\delta_0 \to 0} \delta_0^{\zeta} \ln(\delta_0) = 0 \tag{A.100}$$

for all exponents $\zeta > 0$. □

Estimate of $d^n e^{-\frac{5}{8}ds}$

Lemma A.4.2. *Let the sequences* $(\varepsilon_i)_i$, $(s_i)_i$ *and* $(d_i)_i$ *be as defined in* (4.13). *Under the assumption that*

$$s_0 < \frac{1}{c_1 + c_3 \log(\delta_0^{-1})}, \tag{A.101}$$

where the constants c_1 *and* c_3 *as in* (A.109), *we have that*

$$d_i^n e^{-\frac{5}{8}d_i s_i} < \varepsilon_i, \tag{A.102}$$

for all $i \in \mathbb{N}_0$.

Proof. To begin, we suppress the indexes for readability meaning that we notate s_i as s, d_i as d and so on. By definition of the s and d we have $s^{-1} - 1 < ds < s^{-1}$. Therefore (A.102) is satisfied if

$$d^n e^{-\frac{5}{8}ds} \leq s^{-2n} e^{-\frac{5}{8}\left(s^{-1}-1\right)} = s^{-2n} e^{\frac{5}{8}}\, e^{-\frac{5}{8s}} < \varepsilon \quad \text{or equivalently} \quad s^{-2n} e^{-\frac{5}{8s}} < e^{-\frac{5}{8}}\varepsilon. \tag{A.103}$$

This is in turn implied by

$$s^{-1} + \tfrac{16}{5} n \log(s) > 1 - \tfrac{8}{5}\log\left(\varepsilon\right), \tag{A.104}$$

which is equivalent to

$$s < \left(1 - \tfrac{16}{5} n \log(s) - \tfrac{4}{5}\log(\delta)\right)^{-1} \tag{A.105}$$

We reintroduce the index i, and with the definition of the sequences, we rewrite (A.105) as

$$s_0 < 4^i \left(1 - \tfrac{16}{5} n \log(s_0) + \tfrac{16}{5} n i \log(4) - \tfrac{4}{5}(1+\chi)^i \log(\delta_0)\right)^{-1}, \tag{A.106}$$

or equivalently

$$s_0 < \frac{4^i}{(1+\chi)^i}\left(\frac{1 - \frac{16}{5} n \log(s_0) + \frac{16}{5} n i \log(4)}{(1+\chi)^i} - \tfrac{4}{5}\log(\delta_0)\right)^{-1}. \tag{A.107}$$

Since $\dfrac{4^i}{(1+\chi)^i} \geq 1$, $\dfrac{1}{(1+\chi)^i} \leq 1$, and $\dfrac{i}{(1+\chi)^i} \leq \dfrac{1}{\log(1+\chi)}$ for all $i \in \mathbb{N}$, we have that (A.107) is satisfied for all i if

$$s_0 < \left(1 - \tfrac{16}{5} n \log(s_0) + \frac{\frac{16}{5} n \log(4)}{\log(1+\chi)} - \tfrac{4}{5}\log(\delta_0)\right)^{-1}. \tag{A.108}$$

If we substitute the constants

$$c_1 = 1 + \frac{\frac{16}{5} n \log(4)}{\log(1+\chi)}, \qquad c_2 = \tfrac{16}{5} n, \qquad c_3 = \tfrac{4}{5}, \tag{A.109}$$

we obtain a condition in the form

$$s_0 < \left(c_1 + c_2 \log(s_0^{-1}) + c_3 \log(\delta_0^{-1})\right)^{-1}. \tag{A.110}$$

Lastly, since $c_2 \log(s_0^{-1})$ grows very slow as s_0 decreases, it is usually omitted and compensated by adding a sufficiently large constant to c_1, resulting in the the condition in the form

$$s_0 < \frac{1}{c_1 + c_3 \log(\delta_0^{-1})}, \tag{A.111}$$

for all plausible[2] s_0. This concludes the proof. □

[2]One can pick incredible tiny s_0 such that $c_2 \log(s_0^{-1})$ becomes relevant again, but this has no purpose in KAM theory.

A.5 Tools

This section contains two helpful lemmas, which we used in the other lemmas.

Lemma 1

Lemma A.5.1. *Given $\beta > 0$, there exist a constant $c(\beta) > 0$ such that for all $\sigma > 0$ we have*

$$\sum_{l=1} l^{\beta} e^{-\sigma l} \leq c(\beta)\sigma^{-\beta-1}. \tag{A.112}$$

Proof.

$$\sum_{l=1} l^{\beta} e^{-\sigma l} \leq \int_0^{\infty} \zeta^{\beta} e^{-\sigma\zeta} d\zeta = \sigma^{-\beta-1}\Gamma(\beta+1) \leq c(\beta)\sigma^{-\beta-1}, \tag{A.113}$$

for all $\beta > -1$ and $\sigma > 0$ and where Γ is the Gamma function, which is bounded for finite β. □

This has the simple corollary.

Corollary A.5.2. *For all $\sigma > 0$ we have*

$$\sum_{k\in\mathbb{Z}^n} |k|^{\tau} e^{-\sigma|k|} <\cdot \sigma^{-2\tau-1}. \tag{A.114}$$

Proof.

$$\sum_{k\in\mathbb{Z}^n} |k|^{\tau} e^{-\sigma|k|} \leq 2^n \sum_{l=1}^{\infty} l^{\tau+n-1} e^{-\sigma l} <\cdot \sum_{l=1}^{\infty} l^{2\tau} e^{-\sigma l} <\cdot \sigma^{-2\tau-1}, \tag{A.115}$$

where we used $\#\{k \in \mathbb{Z}^n : |k| = l\} \leq 2^n l^{n-1}$ and lemma A.5.1. □

Lemma 2

Lemma A.5.3. *Let the sequences $(s_i)_i$ and $(d_i)_i$ be defined in* (4.13)*. Under the assumption that*

$$ds > \frac{8}{5}n, \tag{A.116}$$

we have that

$$\sum_{|k|>d_i} e^{-\frac{5}{8}s_i|k|} <\cdot d_i^n e^{-\frac{5}{8}d_i s_i}. \tag{A.117}$$

Proof. To begin, we suppress the index i for readability.

$$\sum_{|k|>d} e^{-\frac{5}{8}s|k|} \leq 2^n \sum_{l=d+1}^{\infty} l^{n-1} e^{-\frac{5}{8}sl} <\cdot \int_d^{\infty} x^{n-1} e^{-\frac{5}{8}sx}\, dx, \tag{A.118}$$

where we used $\#\{k \in \mathbb{Z}^n : |k| = l\} \leq 2^n l^{n-1}$.

$$\int_d^{\infty} x^{n-1} e^{-\frac{5}{8}sx}\, dx = d^{n-1} e^{-\frac{5}{8}sd} \int_d^{\infty} \left(\tfrac{x}{d}\right)^{n-1} e^{-\frac{5}{8}s(x-d)}\, dx \leq d^{n-1} e^{-\frac{5}{8}sd} \int_d^{\infty} \exp\left(\left(\frac{n-1}{d} - \frac{5s}{8}\right)(x-d)\right) dx,$$

where we used that $\left(\frac{x}{d}\right)^{n-1} = e^{(n-1)(\log(x)-\log(d))} \leq \exp\left(\frac{n-1}{d}(x-d)\right)$. We continue with the integral

$$\int_d^{\infty} \exp\left(\left(\frac{n-1}{d} - \frac{5s}{8}\right)(x-d)\right) dx \leq \int_d^{\infty} \exp\left(\frac{-(x-d)}{d}\right) dx = d, \tag{A.119}$$

where we used $\frac{n-1}{d} - \frac{5}{8}s \leq -\frac{1}{d}$, which follows from the assumption (A.116). Combining (A.118), (A.5) and (A.119) results in

$$\sum_{|k|>d} e^{-\frac{5}{8}s|k|} <\cdot d^n e^{-\frac{5}{8}sd}, \tag{A.120}$$

which proves the lemma.

□

Appendix Preliminaries

B.1 Cauchy's estimate

Lemma B.1.1 (Cauchy's estimate for one complex variable)**.** *Let D is an open domain in $\mathbb{C}$, and let $D_r = \{z \in \mathbb{C} : |z - D| < r\}$ be an open neighbourhood of radius r surrounding it. Then let $f : D_r \to \mathbb{C}$ be an analytic function on D_r with a bounded supremum norm $\sup_{z\in D_r} |f(z)| = |f|_{D_r} < \infty$. Then*

$$\left|\tfrac{d^n}{dz^n} f\right|_{D_{r-\rho}} \leq n!\rho^{-n} |f|_{D_r} \tag{B.1}$$

for all $0 < \rho < r$.

Proof. This is a direct consequence of Cauchy's integral formula. □

Cauchy's estimate is easily extended to the multiple dimensional case.

Lemma B.1.2 (Cauchy's estimate for multiple complex variables)**.** *Let D is an open domain in $\mathbb{C}^n$, and let $D_r = \{z \in \mathbb{C}^n : |z - D| < r\}$ be an open neighbourhood of radius r surrounding it. Then let $f : D_r \to \mathbb{C}^n$ be an analytic function on D_r with a bounded supremum norm $\sup_{z\in D_r} |f(z)| = |f|_{D_r} < \infty$. Then*

$$|\partial^\alpha f|_{r-\rho} \leq \alpha!\rho^{-\alpha} |f|_r \tag{B.2}$$

for all $0 < \rho_j < r$, and where we used the multi-index notation for α.

Proof. This directly follows from Cauchy's estimate for one complex variable. □

B.2 Paley-Wiener estimate

Lemma B.2.1. *Let $f : \mathbb{C}^n \to \mathbb{C}^n$ be an analytic function that is bounded in the strip $U_\kappa(T^n)$ and let*

$$f = \sum_{k\in\mathbb{Z}^n} f^k e^{i\langle k,x\rangle} \tag{B.3}$$

be its Fourier series. Then

$$|f^k| \leq |f|_{U_\kappa(T^n)} e^{-|k|s} \tag{B.4}$$

for all $k \in \mathbb{Z}^k$.

For the proof of this lemma see Lemma A.1 in the Lecture notes of Pöschel [73]. This lemma also easily extends to any secondary variable such as y or ω.

B.3 Implicit function theorem

Theorem B.3.1 (Implicit function theorem)**.** *Let $F : \Omega \to \mathbb{R}$ be of class C^1 in an open set Ω inside $\mathbb{R}^n \times \mathbb{R}$ and (a, b) be a point in Ω such that $F(a, b) = 0$ and $\partial_y F(a, b) > 0$.*

Then, there exist an open set X, inside $\mathbb{R}^n$ and containing a, and an open set Y, inside $\mathbb{R}$ and containing b, satisfying the following.

- *For each $x \in X$ there is a unique $y = f(x)$ in Y such that $F(x, f(x)) = 0$.*
- *We have $f(a) = b$. Moreover, $f : X \to Y$ is of class C^1 and*

$$\frac{\partial f}{\partial x_j}(x) = -\frac{\partial_{x_j} F(x, f(x))}{\partial_y F(x, f(x))} \quad \text{for all} \quad x \in X \quad \text{and} \quad j \in \{1, \cdots, n\}. \tag{B.5}$$

For a proof of this theorem consult [64].

B.4 Differential geometry

This section is a very brief list of definitions from differential geometry used in this thesis.

B.4.1 Notation

We introduce a few notations:

- Let d denote the exterior derivative.
- Let i_X denote the interior product with vector field X.
- Let $\mathcal{L}_X$ denote the Lie derivative with respect to vector field X.

B.4.2 Definitions

For convenience the following concepts are reiterated.

Definition B.4.1 (Interior product)**.** A contraction of a differential form with a vector field. Let ω be a p-form, and Y_j vector fields then

$$(i_X\omega)(Y_1, \cdots, Y_{p-1}) := \omega(X, Y_1, \cdots, Y_{p-1}). \tag{B.6}$$

Definition B.4.2 (Closed form)**.** Let ω be a differential form. ω is *closed* if its exterior derivative is 0, i.e. $d\omega = 0$.

Definition B.4.3 (Exact form)**.** Let ω be a differential form. ω is *exact* if it is the exterior derivative of another differential form, i.e. $d\omega^* \equiv \omega$ where ω^* is differential form.

Definition B.4.4 (Lie derivative for differential forms)**.** Let X be a vector field and ω a differential form, then

$$\mathcal{L}_X\omega = d(i_X\omega) + i_X(d\omega). \tag{B.7}$$

This equation is also referred to as Cartan formula or as Cartan's magic formula.

Bibliography

[1] V.I. Arnol'd. "Proof of a theorem of A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian". In: *Russian Math. Surveys* 18.5 (1963), pp. 9–36.

[2] V.I. Arnol'd. "Small divisor problems in classical and Celestial Mechanics". In: *Russian Math. Surveys* 18.6 (1963), pp. 85–191.

[3] V.I. Arnold. "Reversible systems". In: *Nonlinear and Turbulent Processes in Physics* 3 (1984), pp. 1161–1174.

[4] V.I. Arnold and M.B. Sevryuk. "Oscillations and bifurcations in reversible systems". In: *Nonlinear phenomena in plasma physics and hydrodynamics*. Ed. by RZ Sagdeev. Moscow: Mir, 1986, pp. 31–64.

[5] V.I. Bakhtin. "Diophantine approximations on images of mappings." In: *Dokl. Akad. Nauk Beloruss. SSR* 35.5 (1991), pp. 398–400.

[6] Yu.N. Bibikov. "The existence of conditionally periodic solutions of systems of differential equations". In: *Differ. Uravn.* 7.8 (1971), pp. 1347–1356.

[7] Yu.N. Bibikov and V.A. Pliss. "The existence of invariant tori in the neighborhood of the zero solution of a system of ordinary differential equations". In: *Differ. Uravn.* 3.11 (1967), pp. 1864–1881.

[8] A. Bounemoura. "Some remarks on the optimality of the Bruno-Rüssmann condition". In: *Bulletin de la Société mathématique de France* 147.2 (2019), pp. 341–353.

[9] B.L.J. Braaksma and H.W. Broer. "Quasi periodic flow near a codimension one singularity of a divergence free vector field in dimension four". In: *Bifurcation, théorie ergodique et applications: Dijon 1981*. 98-99. Astérisque, 1982, pp. 191–212.

[10] B.L.J. Braaksma, H.W. Broer, and G.B. Huitema. "Unfolding and bifurcations of quasi-periodic tori. II: Toward a quasi-periodic bifurcation theory". In: *Memoirs of the American Mathematical Society* 83 (1990), pp. 83–175.

[11] H. W. Broer, M. Ciocci, and A. Litvak-Hinenzon. "Survey on dissipative KAM theory including quasi-periodic bifurcation theory based on lectures by Henk Broer". In: London Mathematical Society Lecture Note Series 306 (2005). Ed. by Montaldi, James and Ratiu, Tudor, 303–355.

[12] H.W. Broer. "Formal normal form theorems for vector fields and some consequences for bifurcations in the volume preserving case". In: *Dynamical Systems and Turbulence, Warwick 1980*. Ed. by D. Rand and L. Young. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, pp. 54–74. ISBN: 978-3-540-38945-3.

[13] H.W. Broer. "Quasi periodic flow near a codimension one singularity of a divergence free vector field in dimension three". In: *Dynamical Systems and Turbulence, Warwick 1980*. Ed. by D. Rand and L. Young. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, pp. 75–89. ISBN: 978-3-540-38945-3.

[14] H.W. Broer, H. Hanßmann, and J. You. "Bifurcations of normally parabolic tori in Hamiltonian systems". In: *Nonlinearity* 18 (Apr. 2005), pp. 1735–1769. DOI: 10.1088/0951-7715/18/4/018.

[15] H.W. Broer, H. Hanßmann, and J. You. "Umbilical torus bifurcations in Hamiltonian systems". In: *Journal of Differential Equations* 222.1 (2006), pp. 233–262. ISSN: 0022-0396. DOI: https://doi.org/10.1016/j.jde.2005.06.030.

[16] H.W. Broer and G.B. Huitema. "Unfoldings of quasi-periodic tori in reversible systems". In: *Journal of Dynamics and Differential Equations* 7 (1995), pp. 191–212.

[17] H.W. Broer, G.B. Huitema, and M.B. Sevryuk. *Quasi-Periodic Motions in Families of Dynamical Systems*. Vol. 1645. Lecture Notes in Mathematics. Springer Berlin, Heidelberg, 1996. ISBN: 3-540-62025-7.

[18] H.W. Broer, G.B. Huitema, and F. Takens. "Unfolding and bifurcations of quasi-periodic tori. I: Unfolding of quasi-periodic tori". In: *Memoirs of the American Mathematical Society* 83 (1990), pp. 1–82.

[19] H.W. Broer and M.B. Sevryuk. "Chapter 6 - KAM Theory: Quasi-periodicity in Dynamical Systems". In: *Handbook of Dynamical Systems* 3 (Dec. 2010), pp. 249–344. DOI: 10.1016/S1874-575X(10)00314-0.

[20] H.W. Broer and F. Takens. *Dynamical Systems and Chaos*. Vol. 172. Applied Mathematical Sciences. Springer New York, NY, 2011. ISBN: 978-1-4419-6870-8.

[21] A.D. Bruno. "Analytic form of differential equations I". In: *Tr. Mosk. Mat. Obs.* 25 (Jan. 1971), pp. 131–288.

[22] A.D. Bruno. "Analytic form of differential equations II". In: *Tr. Mosk. Mat. Obšc* 26 (Jan. 1972), pp. 199–239.

[23] A.D. Bruno. *Local Methods in Nonlinear Differential Equations*. Berlin, Heidelberg: Springer, 1989.

[24] M. Chaperon. "C^k-conjugacy of holomorphic flows near a singularity". In: *Publications Mathématiques de l'IHÉS* 64 (1986), pp. 143–183.

[25] L. Chierchia. "Kolmogorov's 1954 paper on nearly-integrable Hamiltonian systems". In: *Regul. Chaot. Dyn.* 13 (2008), pp. 130–139.

[26] L. Chierchia. "Kolmogorov-Arnold-Moser (KAM) Theory". In: *Encyclopedia of Complexity and Systems Science*. Ed. by R.A. Meyers. New York: Springer, 2009, pp. 5064–5091.
[27] H. Cremer. "Über die Häufigkeit der Nichtzentren." In: *Math. Ann.* 115 (1938), pp. 573–580.
[28] R. De la Llave. "A tutorial on KAM theory". In: *Smooth Ergodic Theory and Its Applications*. Vol. 69. Proceedings of Symposia in Pure Mathematics. Seattle: Amer. Math. Soc., Jan. 2001, pp. 175–292. ISBN: 9780821826829. DOI: 10.1090/pspum/069/1858536.
[29] H.S. Dumas. *The KAM Story: A Friendly Introduction to the Content, History, and Significance of Classical Kolmogorov–Arnold–Moser Theory*. World Scientific, May 2014. ISBN: 978-9814556583. DOI: 10.1142/8955.
[30] L.C. Evans. "Weak KAM Theory and Partial Differential Equations". In: *Calculus of Variations and Nonlinear Partial Differential Equations: With a historical overview by Elvira Mascolo*. Ed. by Bernard Dacorogna and Paolo Marcellini. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 123–154. ISBN: 978-3-540-75914-0. DOI: 10.1007/978-3-540-75914-0_4.
[31] J. Féjoz. "Démonstration du " théorème d'Arnold " sur la stabilité du système planétaire (d'après Michael Herman)". In: *Ergodic Theory and Dynamical Systems* 24 (Oct. 2004), pp. 1521–1582. DOI: 10.1017/S0143385704000410.
[32] N. Fenichel. "Persistence and Smoothness of Invariant Manifolds for Flows". In: *Indiana University Mathematics Journal* 21.3 (1971), pp. 193–226. ISSN: 00222518, 19435258.
[33] A. Gray. "A fixed point theorem for small divisor problems". In: *Journal of Differential Equations* 18.2 (1975), pp. 346–365. ISSN: 0022-0396. DOI: https://doi.org/10.1016/0022-0396(75)90067-4.
[34] D.M. Grobman. "Homeomorphisms of systems of differential equations. (Translated)". In: *Doklady Akademii Nauk SSSR*. 128 (1959), pp. 880–881.
[35] J. Guckenheimer and P.J. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. New York: Springer-Verlag, 1983.
[36] M. Guzzo. "An Overview on the Nekhoroshev Theorem". In: *Topics in Gravitational Dynamics: Solar, Extra-Solar and Galactic Systems*. Ed. by D. Benest, C. Froeschle, and E. Lega. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 1–28. ISBN: 978-3-540-72984-6. DOI: 10.1007/978-3-540-72984-6_1.
[37] H. Hanßmann. "On Hamiltonian bifurcations of invariant tori with a Floquet multiplier -1". In: *Dynamical Systems* 21.2 (2006), pp. 115–145. DOI: 10.1080/14689360500321440.
[38] H. Hanßmann. *Local and Semi-Local Bifurcations in Hamiltonian Dynamical Systems. Results and Examples*. Vol. 1893. Lecture Notes in Mathematics. Springer Berlin, Heidelberg, 2007. ISBN: 978-3-540-38894-4.
[39] H. Hanßmann. "Non-degeneracy conditions in KAM theory". In: *Indagationes Mathematicae* 22.3 (2011). Devoted to: Floris Takens (1940–2010), pp. 241–256. ISSN: 0019-3577. DOI: https://doi.org/10.1016/j.indag.2011.09.005.
[40] H. Hanßmann. "Quasi-periodic bifurcations in reversible systems". In: *Regular and Chaotic Dynamics* 16 (2011), pp. 51–60.
[41] P. Hartman. "A Lemma in the Theory of Structural Stability of Differential Equations". In: *Proceedings of the American Mathematical Society* 11.4 (1960), pp. 610–620. ISSN: 00029939, 10886826.
[42] P. Hartman. "On the Local Linearization of Differential Equations". In: *Proceedings of the American Mathematical Society* 14.4 (1963), pp. 568–573. ISSN: 00029939, 10886826.
[43] F. Hausdorff. *Set Theory (3rd edition)*. New York: Chelsea Pub. Co., 1991.
[44] M. Hénon. "Explorationes numérique du problème restreint IV: Masses egales, orbites non periodique". In: *Bullettin Astronomique* 3.1 (1996), pp. 49–66.
[45] M.R. Herman. "Recent results and some open questions on Siegel's linearization theorem of germs of complex analytic diffeomorphisms of C^n near a fixed point". In: *VIIIth International Congress on Mathematical Physics (Marseille, 1986)* (1987), pp. 138–184.
[46] M.R. Herman. *Existence et non existence de tores invariants par des difféomorphismes symplectiques*. Séminaire Équations aux dérivées partielles (Polytechnique) dit aussi "Séminaire Goulaouic-Schwartz" (1987-1988), Talk no. 14, 24 p. 1988.
[47] M.R. Herman. "Inégalités a priori pour des tores lagrangiens invariants par des difféomorphismes symplectiques". In: *Publications Mathématiques de l'IHÉS* 70 (1989), pp. 47–101.
[48] M.R. Herman. "Différentiabilité optimale et contre-exemples à la fermeture en topologie C^∞ des orbites récurrentes de flots hamiltoniens". In: *C.R. Acad. Sci. Paris* 313.3 (1991), pp. 49–51.
[49] M.R. Herman. "Exemples de flots hamiltoniens dont aucune perturbation en topologie C^∞ n'a d'orbites périodiques sur un ouvert de surfaces d'énergies". In: *C.R. Acad. Sci. Paris* 313.13 (1991), pp. 989–994.
[50] M.W. Hirsch, C.C. Pugh, and M. Shub. *Invariant Manifolds*. Vol. 583. Lecture Notes in Mathematics. Springer Berlin, Heidelberg, 1977. ISBN: 978-3-540-08148-7.
[51] G.B. Huitema. "Unfoldings of quasi-periodic tori". PhD thesis. Rijksuniversiteit Groningen, 1988.

[52] G.P.X. Koenigs. "Recherches sur les intégrales de certaines équations fonctionnelles". In: *Ann. Sci. École Norm. Sup.* 1 (1884), pp. 2–41.
[53] A. Kolmogorov. "On the conservation of conditionally periodic motions under small perturbations of the Hamiltonian. Dokl Akad Nauk SSSR". In: *Dokl. Akad. Nauk SSSR.* 98 (Jan. 1954), pp. 527–530.
[54] K. Lu, W. Zhang, and W. Zhang. "C^1 Hartman Theorem for random dynamical systems". In: *Advances in Mathematics* 375 (2020), pp. 107–375. ISSN: 0001-8708. DOI: https://doi.org/10.1016/j.aim.2020.107375.
[55] S. Marmi. "Critical functions for complex analytic maps". In: *Journal of Physics A* 23.15 (1990), pp. 3447–3474.
[56] S. Marmi. *An Introduction To Small Divisors*. Sept. 2000. DOI: 10.48550/ARXIV.MATH/0009232.
[57] J.E. Marsden and T.S. Ratiu. "Hamiltonian Systems on Symplectic Manifolds". In: *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*. New York, NY: Springer New York, 1999, pp. 147–164. ISBN: 978-0-387-21792-5. DOI: 10.1007/978-0-387-21792-5_5.
[58] V.K. Melnikov. "On some cases of conservation of conditionally periodic motions under a small change of the Hamiltonian function". In: *Sov. Math. Dokl.* 6 (Jan. 1965), pp. 1592–1596.
[59] J. Moser. "On invariant curves of area-preserving mappings of an annulus". In: *Nachr. Akad. Wiss. Göttingen* 2 (1962), pp. 1–20.
[60] J. Moser. "On the Theory of Quasiperiodic Motions". In: *SIAM Review* 8.2 (1966), pp. 145–172. ISSN: 00361445.
[61] J. Moser. "Convergent series expansions for quasi-periodic motions." In: *Math. Ann.* 169 (1967), pp. 136–176.
[62] J. Moser. *Stable and Random Motions in Dynamical Systems*. Princeton University Press, 1973. ISBN: 9780691089102.
[63] N.N. Nekhoroshev. "An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems". In: *Russian Math. Surveys* 32.6 (1977), pp. 1–65.
[64] O. de Oliveira. "The Implicit and the Inverse Function Theorems: Easy Proofs". In: *Real Analysis Exchange* 39.1 (2014), pp. 207–218.
[65] I.O. Parasyuk. "Conservation of quasiperiodic motions in reversible multifrequency systems". In: *Dokl. Akad. Nauk Ukrain. SSR. Ser. A* 9 (1982), pp. 19–22.
[66] I.O. Parasyuk. "Conservation of multidimensional invariant tori of Hamiltonian systems". In: *Ukrainian Mathematical Journal* 36 (1984), pp. 380–385. DOI: 10.1007/BF01066558.
[67] I.O. Parasyuk. "Coisotropic invariant tori of Hamiltonian systems of the quasiclassical theory of motion of a conduction electron". In: *Ukrainian Mathematical Journal* 42.3 (1990), pp. 308–312. DOI: 10.1007/BF01057014.
[68] I.O. Parasyuk. "Reduction and coisotropic invariant tori of Hamiltonian systems with non-Poisson commutative symmetries. I". In: *Ukrainian Mathematical Journal* 46.4 (1994), pp. 572–580. DOI: 10.1007/BF01058520.
[69] R.P. Pérez-Marco. "Fixed points and circle maps". In: *Acta Math.* 179.2 (1997), pp. 243–294.
[70] G.A. Pfeiffer. "On the Conformal Mapping of Curvilinear Angles. The Functional Equation $\phi[f(x)] = a_1\phi(x)$". In: *Transactions of the American Mathematical Society* 18.2 (1917), pp. 185–198. ISSN: 00029947.
[71] J. Pöschel. "Integrability of Hamiltonian systems on Cantor sets". In: *Communications on Pure and Applied Mathematics* 35 (1982), pp. 653–696.
[72] J. Pöschel. "On Elliptic Lower Dimensional Tori in Hamiltonian Systems." In: *Mathematische Zeitschrift* 202.4 (1989), pp. 559–608.
[73] J. Pöschel. "A Lecture on the Classical KAM-Theorem". In: *Proc. Symp. Pure Math.* 69 (Dec. 2001), pp. 707–732. DOI: 10.1090/pspum/069/1858551.
[74] C.C. Pugh. "On a Theorem of P. Hartman". In: *American Journal of Mathematics* 91.2 (1969), pp. 363–367. ISSN: 00029327, 10806377.
[75] A.S. Pyartli. "Diophantine approximations on submanifolds of Euclidean space." In: *Funct. Anal. Its. Appl.* 3 (1969), pp. 303–306.
[76] H.R. Rüssmann. "Über die Iteration analytischer Funktionen". In: *Journal of Mathematics and Mechanics* 17.6 (1967), pp. 523–532. ISSN: 00959057, 19435274.
[77] H.R. 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: Battelle Seattle 1974 Rencontres*. Ed. by J. Moser. Berlin, Heidelberg: Springer Berlin Heidelberg, 1975, pp. 598–624. ISBN: 978-3-540-37505-0. DOI: 10.1007/3-540-07171-7_12.
[78] H.R. Rüssmann. "On the one-dimensional Schrodinger equation with a quasi-periodic potential". In: *Nonlinear Dynamics* 357.1 (Dec. 1980), pp. 90–107.
[79] H.R. Rüssmann. "Non-degeneracy in the perturbation theory of integrable dynamical systems". In: *London Mathematical Society Lecture Note Series* 134 (1989), pp. 5–18.

[80] H.R. Rüssmann. "On the Frequencies of Quasi Periodic Solutions of Analytic Nearly Integrable Hamiltonian Systems". In: *Seminar on Dynamical Systems: Euler International Mathematical Institute, St. Petersburg, 1991*. Ed. by S. Kuksin, V. Lazutkin, and J. Pöschel. Basel: Birkhäuser Basel, 1994, pp. 160–183. ISBN: 978-3-0348-7515-8. DOI: 10.1007/978-3-0348-7515-8_13.

[81] H.R. Rüssmann. "Invariant Tori in Non-Degenerate Nearly Integrable Hamiltonian Systems". In: *Regul. Chaotic Dyn.* 6.2 (2001), pp. 119–204.

[82] J. Scheurle. "A class of quasiperiodic solutions for the plane three-body problem". In: *Methoden Verfahren Math. Phys. Frankfurt a. M. (Proc. Conf. Oberwolfach, 1980)*. Vol. 21. Peter D. Lang Verlag, 1981, pp. 71–80.

[83] J. Scheurle. "Quasiperiodic solutions of the plane three-body problem near Euler's orbits". In: *Celestial Mechanics* 28.1-2 (1982), pp. 141–151.

[84] J. Scheurle. "Bifurcation of quasiperiodic solutions from equilibrium points of reversible dynamical systems". In: *Archive for Rational Mechanics and Analysis* 97.2 (1987), pp. 103–139.

[85] E. Schröder. "Über iterierte Funktionen". In: *Math. Ann.* 3 (1871), pp. 296–322.

[86] M.B. Sevryuk. *Reversible Systems*. Vol. 1211. Lecture Notes in Mathematics. Springer Berlin, Heidelberg, 1986. ISBN: 978-3-540-16819-5.

[87] M.B. Sevryuk. "The classical KAM theory at the dawn of the twenty-first century". In: *Mosc. Math. J.* 3.3 (2003), pp. 1113–1144.

[88] M.B. Sevryuk. "The reversible context 2 in KAM theory: The first steps". In: *Regular and Chaotic Dynamics* 16 (Mar. 2011), pp. 24–38. DOI: 10.1134/S1560354710520035.

[89] M.B. Sevryuk. "KAM theory for lower dimensional tori within the reversible context 2". In: *Mosc. Math. J.* 12.2 (2012), pp. 435–455.

[90] M.B. Sevryuk. "Quasi-periodic perturbations within the reversible context 2 in KAM theory". In: *Indagationes Mathematicae* 23.3 (2012), pp. 137–150. ISSN: 0019-3577. DOI: https://doi.org/10.1016/j.indag.2011.12.004.

[91] M.B. Sevryuk. "Hamiltonian and reversible systems with smooth families of invariant tori". In: *Indagationes Mathematicae* 32.2 (2021), pp. 406–425. ISSN: 0019-3577. DOI: https://doi.org/10.1016/j.indag.2020.12.001.

[92] C.L. Siegel. "Iteration of analytical functions". In: *Ann. Math.* 43 (1942), pp. 265–269.

[93] C.L. Siegel. *Gesammelte Abhandelungen*. Berlin, Heidelberg: Springer Verslag, 1966.

[94] E.M. Stein. *Singular Integrals and Differentiability Properties of Functions (PMS-30)*. Princeton University Press, 1970. ISBN: 9780691080796.

[95] S. Sternberg. "Local Contractions and a Theorem of Poincare". In: *American Journal of Mathematics* 79.4 (1957), pp. 809–824. ISSN: 00029327, 10806377.

[96] S. Sternberg. "On the Structure of Local Homeomorphisms of Euclidean n-Space, II". In: *American Journal of Mathematics* 80.3 (1958), pp. 623–631.

[97] S. Sternberg. "Infinite Lie Groups and the Formal Aspects of Dynamical Systems". In: *Journal of Mathematics and Mechanics* 10.3 (1961), pp. 451–474. ISSN: 00959057, 19435274.

[98] L.W. Tu. *An Introduction to Manifolds*. Universitext. Springer New York, 2010. ISBN: 9781441973993. URL: https://books.google.nl/books?id=br1KngEACAAJ.

[99] H. Whitney. "Analytic Extensions of Differentiable Functions Defined in Closed Sets". In: *Transactions of the American Mathematical Society* 36.1 (1934), pp. 63–89.

[100] J.C. Yoccoz. "Théorème de Siegel, nombres de Bruno et polynômes quadratiques". In: *Astérisque* 231 (1995), pp. 3–88.

[101] J.C. Yoccoz. "Linéarisation des germes de difféomorphismes holomorphes de $(C, 0)$." In: *Comptes rendus de l'Académie des Sciences* 306.1 (1999), pp. 55–58.

[102] Y. Yu, J. Ye, W. Liu, and C. Zhang. "Quantum analog of the Kolmogorov-Arnold-Moser theorem in the anisotropic Dicke model and its possible implications in the hybrid Sachdev-Ye-Kitaev models". In: *Phys. Rev. A* 106.2 (Aug. 2022), p. 022213. DOI: 10.1103/PhysRevA.106.022213.

[103] E. Zehnder. "Generalized implicit function theorems with applications to some small divisor problems, I". In: *Communications on Pure and Applied Mathematics* 28.1 (1975), pp. 91–140.

[104] E. Zehnder. "A simple proof of a generalization of a theorem by C. L. Siegel". In: *Lecture notes in Math.* 597 (1977), pp. 855–866.