Higher-order averaging, formal series and numerical integration II: the quasi-periodic case

Ph. Chartier,* A. Murua,† J. M. Sanz-Serna‡

October 8, 2015

Abstract

The paper considers non-autonomous oscillatory systems of ordinary differential equations with $d \geq 1$ non-resonant constant frequencies $\omega_1, \ldots, \omega_d$. Formal series like those used nowadays to analyze the properties of numerical integrators are employed to construct higher-order averaged systems and the required changes of variables. With the new approach, the averaged system and the change of variables consist of vector-valued functions that may be written down immediately and scalar *coefficients* that are universal in the sense that they do not depend on the specific system being averaged and may therefore be computed once and for all given $\omega_1, \ldots, \omega_d$. The new method may be applied to obtain a variety of averaged systems. In particular we study the *quasi-stroboscopic* averaged system characterized by the property that the true oscillatory solution and the averaged solution coincide at the initial time. We show that quasi-stroboscopic averaging is a geometric procedure because it is independent of the particular choice of co-ordinates used to write the given system. As a consequence, quasi-stroboscopic averaging of a canonical Hamiltonian (resp. of a divergence-free) system results in a canonical (resp. in a divergence-free) averaged system. We also study the averaging of a family of near-integrable systems where our approach may be used to construct explicitly d formal first integrals for both the given system and its quasi-stroboscopic averaged version. As an application we construct three first integrals of a system that arises as a nonlinear perturbation of five coupled harmonic oscillators with one slow frequency and four resonant fast frequencies.

Communicated by Ernst Hairer

Keywords and sentences: Averaging, high-order averaging, quasi-stroboscopic averaging, highly oscillatory problems, Hamiltonian problems, formal series, B-series, trees, first-integrals, near-integrable systems, Lie groups, Lie algebras

Mathematics Subject Classification (2010) 34C29, 65L06, 34D20, 70H05

1 Introduction

The aim of this series of papers is to show how the formal series expansions that are nowadays used to analyze numerical integrators [13], [14], [18], [23], provide a powerful means to study and implement the method of averaging (see e.g. [17], [22] and also [1], Chapter 4, [2], Chapter 10). While Part I of the series [10] was restricted to the case of systems with a single fast frequency, in this paper —which may be read independently from [10]— we

*INRIA Rennes and ENS Cachan Bretagne, Campus Ker-Lann, av. Robert Schumann, 35170 Bruz, France. Email: Philippe.Chartier@inria.fr

†Konputazio Zientziak eta A. A. Saila, Informatika Fakultatea, UPV/EHU, E–20018 Donostia–San Sebastián, Spain. Email: Ander.Murua@ehu.es

‡(Corresponding author) Departamento de Matemática Aplicada, Facultad de Ciencias, Universidad de Valladolid, Valladolid, Spain. Email: sanzsern@mac.uva.es

consider systems with $d \geq 1$ constant fast frequencies. The approach described here makes it possible to construct in an *explicit* way higher-order averaged systems and the associated changes of variables. Before summarizing the contributions of the paper in Subsection 1.2 below, we shall briefly review in Subsection 1.1 those basic aspects of the method of higher-order averaging that are essential to outline the scope of our research.

1.1 High-order averaging in quasi-periodic systems

We consider averaging as applied to an initial value problem (see e.g. [21], [24])

$$\frac{d}{dt}y = \varepsilon f(y, t\omega), \tag{1}$$

$$y(0) = y_0 \in \mathbb{R}^D, \tag{2}$$

in an interval $0 \leq t \leq L/\varepsilon$ whose length increases as $\varepsilon \to 0$. Here $f = f(y, \theta)$ depends 2π-periodically on each of the scalar components $\theta_1,\ldots,\theta_d$ of the angular variable $\theta \in \mathbb{T}^d$ and $\omega \in \mathbb{R}^d$ is a constant vector of angular frequencies. Throughout the paper we assume that ω is *non-resonant*, i.e. $\mathbf{k} \cdot \omega \neq 0$ for each multi-index $\mathbf{k} \in \mathbb{Z}^d$, with $\mathbf{k} \neq \mathbf{0}$.[1] The case where $f = f(y, \theta; \varepsilon)$ depends explicitly on ε may be reduced to the format (1), see e.g. [21].[2]

In N-th order averaging, $N = 1, 2, \ldots$, the solution y of (1)–(2) is represented as

$$y(t) = U_N(Y_N(t), t\omega) + \mathcal{O}(\varepsilon^N), \tag{3}$$

where, in the simplest case,[3] U_N is a change of variables of the form

$$U_N(Y, \theta) = Y + \varepsilon u_1(Y, \theta) + \cdots + \varepsilon^{N-1} u_{N-1}(Y, \theta), \qquad \theta \in \mathbb{T}^d, \tag{4}$$

and $Y_N(t)$ is the solution of an *autonomous* problem (the averaged problem)

$$\frac{d}{dt}Y = \varepsilon F_1(Y) + \varepsilon^2 F_2(Y) + \cdots + \varepsilon^N F_N(Y), \qquad Y(0) = \xi, \tag{5}$$

whose initial value is determined ξ implicitly through the equation $U_N(\xi, \mathbf{0}) = y_0$, i.e.

$$\xi + \sum_{j=1}^{N-1} \varepsilon^j u_j(\xi, \mathbf{0}) = y_0. \tag{6}$$

The representation (3) implies that, except for the $\mathcal{O}(\varepsilon^N)$ remainder, the behaviour of y is determined by the behavior of the averaged solution Y_N in tandem with the quasi-periodic rotation $t \mapsto t\omega$.

The functions F_j and u_j may be defined recursively by ($j \geq 1$):

$$F_j(Y) := \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} \tilde{F}_j(Y, \theta) d\theta, \tag{7}$$

$$\omega \cdot \nabla_\theta u_j(Y, \theta) = \tilde{F}_j(Y, \theta) - F_j(Y), \tag{8}$$

$$\int_{\mathbb{T}^d} u_j(Y, \theta) d\theta = 0, \tag{9}$$

where

$$\tilde{F}_1(Y, \theta) := f(Y, \theta)$$

[1] Of course a problem with a resonant ω may be written in non-resonant form by lowering the number d of frequencies.

[2] Alternatively it is also possible to let f depend smoothly on ε and expand in powers of ε as in [10].

[3] In practice the change of variables is often expressed in alternative ways, for instance by using Lie series. With such alternatives, the recursion (7)–(10) has to be changed accordingly. However the exact format of the change of variables and the recursion are not essential to the discussion that follows and we choose to focus on the form (4) for the sake of being definite. See in this connection Remark 2.16 that shows the relations between different averaged systems.

and, for $j > 1$,

$$\tilde{F}_j(Y,\theta) := \sum_{r=1}^{j-1} \Big[\frac{1}{r!} \sum_{i_1+\cdots+i_r=j-1} \frac{\partial^r f}{\partial y^r}(Y,\theta)\Big(u_{i_1}(Y,\theta),\ldots,u_{i_r}(Y,\theta)\Big) - \frac{\partial u_r}{\partial Y}(Y,\theta)F_{j-r}(Y)\Big], \qquad (10)$$

In (10), $(\partial^r f/\partial y^r)(y,\theta)$ denotes the r-th Fréchet derivative of f with respect to y evaluated at (y,θ), a multi-linear function that maps $\mathbb{R}^D \times \overset{r}{\cdots} \times \mathbb{R}^D$ into $\mathbb{R}^D$. The effective computation of the F_j's and u_j's from the relations (7)–(10) is of course a formidable task even for moderate values of j and computer algebra tools are very useful in this respect.

It is well known that, except in the case $d = 1$ of a single frequency, the phenomenon of small denominators introduces substantial difficulties in the derivation of estimates for the remainder in (3) and in fact makes it necessary to strengthen the hypotheses on ω beyond the requirement of non-resonance. We quote the following result from [21], probably the oldest reference (the notation is as above):

Theorem 1.1 (Perko) *Consider the system (1)–(2) where $f(y,\theta)$ is real-analytic and 2π-periodic in each component of θ for each y in a convex region G (containing y_0) and has continuous partial derivatives of order N with respect to y in $G \times \mathbb{T}^d$. Assume that the vector $\omega \in \mathbb{R}^d$ satisfies the* strong non-resonance *condition*

$$\forall \mathbf{k} \in \mathbb{Z}^d \backslash \{\mathbf{0}\}, \qquad |\mathbf{k}\cdot\omega| \geq c|\mathbf{k}|^{-\nu}$$

for some positive constants c and ν. Assume in addition that the solution of the initial value problem

$$\frac{d}{dt}Y = \varepsilon F_1(Y), \qquad Y(0) = y_0,$$

has a solution $Y_1(t)$ which remains in G for $0 \leq t \leq L/\varepsilon$. Then, there exist $\varepsilon_N, C_N > 0$ such that, for $0 < \varepsilon < \varepsilon_N$, the problem (1)–(2) has a unique solution in G for $0 \leq t \leq L/\varepsilon$ that satisfies

$$\|y(t) - U(Y_N(t),\omega t)\| \leq C_N \varepsilon^N.$$

Note that, for fixed Y, (8) is a differential equation (the *homological equation*) where the unknown function u_j has to depend periodically on θ; straightforward Fourier analysis shows that the equation is solvable because $\tilde{F}_j - F_j$ has zero angular average. However the angular average of u_j is not determined by (8) and an additional condition is needed. Above that condition is given by (9) which is equivalent to the requirement that the angular average of the function $U_N(Y,\theta)$ in (4) coincides with Y or, in other words, the change of variables is the *identity on average*. There are alternative ways of performing high-order averaging; in them (9) is replaced by a suitable condition that, together with (8), determines u_j. For instance, in the case $d = 1$ of a single frequency $\omega = \omega_1$, it is possible to replace (9) by the condition $u_j(Y,0) = 0$; then $U_N(Y,\theta)$ in (4) coincides with Y at $\theta_1 = 0$ and therefore the change of variables is the *identity at the initial time* $t = 0$. With this alternative there is no need to solve an equation to find the initial value ξ in (6); one simply has $\xi = y_0$. Furthermore, due to the periodicity, (4) implies that $y(t)$ and $Y_N(t)$ also coincide at the *stroboscopic times* $n(2\pi/\omega_1)$, n integer, and accordingly the technique is known as *stroboscopic averaging* [22].

1.2 Scope

In this paper we show how the use of B-series [15] allows the *effective* explicit computation of the u_j's and F_j's in (4) and (5) bypassing the recursive procedure in (7)–(10). With B-series, each u_j and F_j is written by combining so-called elementary differentials and coefficients. The elementary differentials are functions constructed systematically in a very simple way from the function f in (1). The coefficients are numbers that are *universal*, in the sense that they do not depend on the specific choice of D, f or y_0, and therefore, given ω, they may be computed once and for all. The representation in terms of coefficients and elementary differentials has proved to be very helpful to analyze standard numerical methods for differential equations and also to compare such methods,

because it separates the effects of the integrator (incorporated into the coefficients) from those of the system being integrated (which determines the elementary differentials). The same consideration may apply here to integrators for highly-oscillatory problems; modulated Fourier expansions [12] are of course another similar tool.

The new technique may be applied to perform averaging not only as in (7)–(10), but also through variants where the u_j's are not determined by imposing zero angular average by means of (9). In particular the B-series approach leads in a natural way to a generalization of stroboscopic averaging to the quasi-periodic, $d > 1$, situation; this generalization, where in (5) $\xi = y_0$, will be referred to as *quasi-stroboscopic averaging*.

The new approach may also be advantageously applied to the derivation of estimates for the remainder in (3); such estimates (that may be exponentially small as in [20] or [24]) are the subject of [11] and will not be considered further here. Since we shall not be dealing with the remainder term, it is convenient to replace (4) and (5) in what follows by

$$U(Y,\theta) = Y + \varepsilon u_1(Y,\theta) + \varepsilon^2 u_2(Y,\theta) + \cdots ,$$

and

$$\frac{d}{dt}Y = \varepsilon F_1(Y) + \varepsilon^2 F_2(Y) + \cdots ,$$

respectively. Truncation of these formal series in powers of ε gives the sought formulae (4) and (5) for any required value of N.

The main ideas of the paper are contained in Section 2. To make the article more accessible, we begin by presenting in Subsections 2.1–2.3 an introduction to rooted trees and B-series, the basic tools used in the sequel. Subsection 2.3 ends (see (21)) by showing how to rewrite the given problem (1)–(2) as an initial value problem for a set of complex-valued coefficients $\alpha_u(t)$ that is completely independent of the vector field f. Subsection 2.4 is devoted to investigating the properties of these coefficients; in particular we show their relation to a *transport partial differential equation* (see (23)) which plays here the role played by the homological equation (8) in conventional treatments. After these preliminaries, the technique of quasi-stroboscopic averaging is described in Subsection 2.5 (see Theorem 2.10). When the number d of frequencies is 1, quasi-stroboscopic averaging reduces to the stroboscopic averaging method investigated in Part 1[4] and therefore our results here are an extension of those in [10]; however the derivation in this paper is simpler and more direct than that provided in [10]. Subsection 2.6 considers other possible averaged systems beyond the one obtained via quasi-stroboscopic averaging; we show that all such averaged systems may in fact be derived by first performing stroboscopic averaging and then changing variables. Subsection 2.7 studies the effect on the averaged system of prescribing the initial condition (2) at a time $t_0 \neq 0$.

Section 3 is devoted to the geometric properties of averaging. It turns out that quasi-stroboscopic averaging is a geometric procedure in the sense that it commutes with changes of variables. As a consequence, the quasi-stroboscopic averaged system may be written (see (60)) in terms of commutators (Lie brackets) of the Fourier coefficients of the field f in (1). This in turn implies that if the original system is divergence-free or canonical (Hamiltonian) the quasi-stroboscopic averaged system will also be divergence-free or canonical.

In Section 4 we provide a simple recursion to compute the coefficients of the quasi-stroboscopic averaged systems. By combining the coefficient values obtained from that recursion with the rewriting in terms of commutators linked to the geometric properties of Section 3, we find a compact general expression for the quasi-stroboscopic averaged system with an $\mathcal{O}(\varepsilon^4)$ error term (see (64)).

Section 5 considers a family of autonomous problems where the vector field is an $\mathcal{O}(\varepsilon)$ perturbation of an integrable system. Such problems may be brought to the format (1) by means of a time-dependent change of variables. We describe how to obtain a quasi-stroboscopically averaged system (Theorem 5.4) with a number of favorable properties (Theorem 5.5). If the original system is Hamiltonian, the averaged system will also be Hamiltonian (Theorem 5.8) and furthermore it is possible to construct explicitly d formal first integrals of both the given and averaged systems. As an application we construct three first integrals of a system taken from [13] that arises as a nonlinear perturbation of five coupled harmonic oscillators with one slow frequency and four resonant fast frequencies.

There is a short Appendix that illustrates the use of the product $*$ of maps on trees.

[4] For the application of the idea of stroboscopic averaging to the construction of numerical integrators the reader is referred to [5], [6].

2 Averaging via B-series

In what follows we assume that $f(y,\theta)$ in (1) has continuous partial derivatives of all orders with respect to y in $\mathbb{R}^D \times \mathbb{T}^d$. We denote by $f_{\mathbf{k}}(y)$ the Fourier coefficients of f, so that the corresponding Fourier expansion is

$$f(y,\theta) = \sum_{\mathbf{k}\in\mathbb{Z}^d} e^{i\mathbf{k}\cdot\theta} f_{\mathbf{k}}(y); \tag{11}$$

the $f_{\mathbf{k}}(y)$ are complex functions, but, in order to have a real system, it is necessary that, for each $\mathbf{k}$, $f_{\mathbf{k}} \equiv f^*_{-\mathbf{k}}$, where * denotes complex conjugate.[5] We shall furthermore assume that, except for a finite number of values of $\mathbf{k} \in \mathbb{Z}^d$, all the $f_{\mathbf{k}}$'s vanish identically.

We have placed stringent hypotheses on f so as not to clutter the presentation with unwelcome details. In fact it is possible to replace $\mathbb{R}^D$ by a domain $G \subset \mathbb{R}^D$ and, if (4) and (5) are only of interest for a particular value of N, continuous partial derivatives of order N with respect to y are sufficient. The hypothesis that the Fourier expansion consists of a finite number of (nontrivial) terms is only used to ensure the convergence of a number of series that will appear later (see Remark 2.2). Such a convergence may be ensured by imposing the alternative assumptions, used in Theorem 1.1, that f depends analytically on θ and that ω is strongly non-resonant. Finally we note that since most results in this paper refer to the coefficients in the expansions and these are universal, i.e. independent of f, the hypotheses on f play no real role in most developments.

2.1 The expansion of the oscillatory solution

While formulae (7)–(10) are derived by studying the way in which changes of variables transform an oscillatory differential system, our approach starts by studying the nature of the oscillatory solution y. In order to motivate the combinatorial and algebraic material required later, we obtain in this subsection the first few terms of the formal expansion in powers of ε of the solution y of the initial-value problem (1)–(2), with f given by (11). The problem may be rewritten in integral form as

$$y(t) = y_0 + \varepsilon \int_0^t f(y(s), s\omega)\, ds = y_0 + \varepsilon \int_0^t \sum_{\mathbf{k}\in\mathbb{Z}^d} e^{is\mathbf{k}\cdot\omega} f_{\mathbf{k}}(y(s))\, ds, \tag{12}$$

and therefore $y(t) = y_0 + \mathcal{O}(\varepsilon)$ (in the sense of formal power series). Inserting *à la Picard* this approximation into (12) we write

$$y(t) = y_0 + \varepsilon \sum_{\mathbf{k}\in\mathbb{Z}^d} \left(\int_0^t e^{is\mathbf{k}\cdot\omega}\, ds \right) f_{\mathbf{k}}(y_0) + \mathcal{O}(\varepsilon^2),$$

an expression that may be taken again to (12) to obtain, after Taylor expansion,

$$\begin{aligned} y(t) \;=\; & y_0 + \varepsilon \sum_{\mathbf{k}\in\mathbb{Z}^d} \left(\int_0^t e^{is\mathbf{k}\cdot\omega}\, ds \right) f_{\mathbf{k}}(y_0) \\ & + \varepsilon^2 \sum_{\mathbf{k},\mathbf{l}\in\mathbb{Z}^d} \left(\int_0^t \int_0^{s_1} e^{i(s_1\mathbf{k}+s_2\mathbf{l})\cdot\omega}\, ds_1 ds_2 \right) f'_{\mathbf{k}}(y_0) f_{\mathbf{l}}(y_0) + \mathcal{O}(\varepsilon^3), \end{aligned} \tag{13}$$

where $f'_{\mathbf{k}}(y_0)$ denotes the (Fréchet) derivative (Jacobian matrix) of $f_{\mathbf{k}}$ evaluated at y_0. This procedure may clearly be iterated to find successively the ε^r, $r = 3, 4, \ldots$, terms of the expansion of $y(t)$. The complexity of the corresponding expressions increases rapidly with r and the mode-colored rooted trees to be introduced in the next subsection provide a means whereby the expansion may be found in a systematic way.

It is important to emphasize at this stage that (13) includes elements of two kinds:

[5] We work with real systems for the sake of being definite; all our results are valid for complex systems.

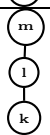
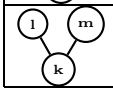
u	$\mathcal{F}_u(y)$	$\alpha_u(t)$	σ_u
	$f_{\mathbf{k}}(y)$	$\int_0^t e^{is_1\mathbf{k}\cdot\omega}\,ds_1$	1
	$f'_{\mathbf{k}}(y)f_{\mathbf{l}}(y)$	$\int_0^t\int_0^{s_1} e^{i(s_1\mathbf{k}+s_2\mathbf{l})\cdot\omega}\,ds_1\,ds_2$	1
	$f'_{\mathbf{k}}(y)f'_{\mathbf{l}}(y)f_{\mathbf{m}}(y)$	$\int_0^t\int_0^{s_1}\int_0^{s_2} e^{i(s_1\mathbf{k}+s_2\mathbf{l}+s_3\mathbf{m})\cdot\omega}\,ds_1\,ds_2\,ds_3$	1
	$f''_{\mathbf{k}}(y)(f_{\mathbf{l}}(y),f_{\mathbf{m}}(y))$	$\int_0^t e^{is_1\mathbf{k}\cdot\omega}\Big(\int_0^{s_1} e^{is_2\mathbf{l}\cdot\omega}\,ds_2\int_0^{s_1} e^{is_3\mathbf{m}\cdot\omega}\,ds_3\Big)ds_1$	$1+\delta(\mathbf{l},\mathbf{m})$

Table 1: Families of trees of orders ≤ 3 with their associated elementary differentials, coefficients and symmetries. Here δ represents Kronecker's symbol, i.e. $\delta(\mathbf{l},\mathbf{m})=1$ if $\mathbf{l}=\mathbf{m}$ and $\delta(\mathbf{l},\mathbf{m})=0$ if $\mathbf{l}\neq\mathbf{m}$

1. *Elementary differentials*, i.e. vector-valued expressions $f_{\mathbf{k}}(y_0)$, $f'_{\mathbf{k}}(y_0)f_{\mathbf{l}}(y_0)$ that depend on the initial-value problem (1)–(2) under consideration,
2. Scalar *coefficients*
$$\int_0^t e^{is\mathbf{k}\cdot\omega}\,ds, \qquad \int_0^t\int_0^{s_1} e^{i(s_1\mathbf{k}+s_2\mathbf{l})\cdot\omega}\,ds_1ds_2,$$
that are *universal*, i.e. independent of the specific choice of D, f and y_0 in (1)–(2).

2.2 Mode-coloured rooted trees and their elementary differentials

We consider rooted trees where each vertex has been colored (i.e. labelled) with a multi-index $\mathbf{k}\in\mathbb{Z}^d$ which refers to a Fourier mode in the expansion (11). For simplicity we shall use hereafter the word *tree* to refer to such a *mode-coloured rooted tree.* The left-most column of Table 1 lists the four families of trees with ≤ 3 vertices. Each family contains infinitely-many trees (for instance there is a tree with one vertex for each particular value of the multi-index $\mathbf{k}$). In the table the lowest vertex of each graph corresponds to the root of the tree.

Formally the set $\mathcal{T}$ of trees may be defined recursively by the following two rules:

1. For each $\mathbf{k}\in\mathbb{Z}^d$, the tree (with one vertex —the root— colored by $\mathbf{k}$) belongs to $\mathcal{T}$.
2. If $u_1,\ldots,u_n$ are trees[6] $\in\mathcal{T}$, then the tree
$$u=[u_1\cdots u_n]_{\mathbf{k}}\,, \tag{14}$$
obtained by connecting through an edge their roots to a new root labelled with the multi-index $\mathbf{k}\in\mathbb{Z}^d$, belongs to $\mathcal{T}$.

With each tree $u\in\mathcal{T}$ we associate two integers: the *order* $|u|$ is the number of vertices of u and the *symmetry* σ_u is defined recursively as follows:

1. For each tree of order 1, $\sigma_u:=1$.

[6]Repetitions among the u_ν are allowed. The order of the u_ν's is of no consequence: thus $[u_1,u_2]_{\mathbf{k}}$ and $[u_2,u_1]_{\mathbf{k}}$ are the same tree even if $u_1\neq u_2$.

2. For u in (14),
$$\sigma_u := r_1! \cdots r_m!\, \sigma_{u_1}^{r_1} \cdots \sigma_{u_m}^{r_m},$$
where u_μ, $\mu = 1, \ldots, m$ denote the pairwise distinct u_ν, $\nu = 1, \ldots, n$ and r_μ counts the number of times that u_μ appears among the u_ν.

The integer σ_u is higher for trees that have a more symmetric appearance: the tree in the last row of Table 1 has $\sigma_u = 2$ if $\mathbf{l} = \mathbf{m}$ and $\sigma_u = 1$ in other cases.

Finally, with each tree $u \in \mathcal{T}$ we associate its *elementary differential* (relative to the functions $f_{\mathbf{k}}$ in (11)) a map $\mathcal{F}_u : \mathbb{R}^D \to \mathbb{R}^D$ defined by:

1. For each tree $\textcircled{\scriptsize k}$ of order 1, $\mathcal{F}_u(y) := f_{\mathbf{k}}(y)$.
2. For u in (14),
$$\mathcal{F}_u(y) := \frac{\partial^n f_{\mathbf{k}}}{\partial y^n}(y)\Big(\mathcal{F}_{u_1}(y), \ldots, \mathcal{F}_{u_n}(y)\Big).$$

We emphasize that even though this definition is recursive, it is a trivial matter to write down, *without following any recursive procedure,* the elementary differential that corresponds to any given tree. To each vertex in the tree there corresponds a Fréchet derivative in the elementary differential; the derivative is the n-th if the vertex has n children. The label $\mathbf{k}$ of a vertex determines the function $f_{\mathbf{k}}$ to be differentiated (see the first two columns in Table 1).

With these definitions we look for an expansion of the form (cf. (13))
$$y(t) = y_0 + \sum_{r=1}^{\infty} \varepsilon^r \sum_{|u|=r} \frac{1}{\sigma_u} \alpha_u(t) \mathcal{F}_u(y_0) = y_0 + \sum_{u \in \mathcal{T}} \frac{\varepsilon^{|u|}}{\sigma_u} \alpha_u(t) \mathcal{F}_u(y_0), \tag{15}$$
where the values of the scalar *coefficients* $\alpha_u(t)$ have to be determined (the introduction of the factor $1/\sigma_u$ here simplifies later formulae). Substitution of this *ansatz* for $y(t)$ into (12) leads, after suitable Taylor expansions, to the conditions
$$\forall \mathbf{k} \in \mathbb{Z}^d, \qquad \alpha_{\textcircled{\scriptsize k}}(t) = \int_0^t e^{is\mathbf{k}\cdot\omega}\, ds, \tag{16}$$
$$\forall u = [u_1 \cdots u_n]_{\mathbf{k}}, \qquad \alpha_u(t) = \int_0^t e^{is\mathbf{k}\cdot\omega} \alpha_{u_1}(s) \cdots \alpha_{u_n}(s)\, ds. \tag{17}$$
Since the trees in the right hand-side of (17) have order $< |u|$, this formula makes it possible to compute recursively the coefficients α_u (see the third column in Table 1).[7] We sum up our findings so far in the following proposition:

Proposition 2.1 *The solution $y(t)$ of the initial-value problem (1)–(2), with f given by (11), possesses the formal expansion in powers of ε given by (15), where the coefficients $\alpha_u(t)$ are recursively defined in (16)–(17).*

2.3 B-series

In this subsection we introduce the formal series that are the main tool used in this paper. If δ is a map that associates with each $u \in \mathcal{T} \cup \{\emptyset\}$ a complex number δ_u, the *B-series* defined by δ is:
$$B(\delta, y) := \delta_\emptyset y + \sum_{r=1}^{\infty} \varepsilon^r \sum_{|u|=r} \frac{1}{\sigma_u} \delta_u \mathcal{F}_u(y) = \delta_\emptyset y + \sum_{u \in \mathcal{T}} \frac{\varepsilon^{|u|}}{\sigma_u} \delta_u \mathcal{F}_u(y). \tag{18}$$
Note that the B-series depends, through the elementary differentials $\mathcal{F}_u$, on the choice of the functions $f_{\mathbf{k}}$ in (11), even though this dependence has not been incorporated into the notation. According to (15), for each fixed t, the expansion for $y(t)$ is the B-series $B(\alpha(t), y_0)$, where $\alpha_\emptyset(t) = 1$ and for $u \in \mathcal{T}$ the coefficient $\alpha_u(t)$ is defined through (16)–(17).

[7] The effective computation of these coefficients will be considered in Section 4.

Remark 2.2 The hypotheses on f at the beginning of this section ensure that for each r the inner (infinite) series in (18) consists of a finite number of nontrivial terms and therefore is a well-defined function of y; as a result $B(\delta, y)$ is a well-defined formal series in powers of ε. As mentioned before, for strongly non-resonant ω, the hypothesis that f consists of finitely-many Fourier modes may be relaxed to analytic dependence of f on θ as in Perko's Theorem 1.1. In such a relaxed scenario, one has to build on the decay of $f_{\mathbf{k}}$ as $|\mathbf{k}| \to \infty$ to ensure that, for well-behaved maps δ, the inner series in (18) is a well-defined function of y. See [11] for further discussion of this point.

Remark 2.3 All formulae below involving B-series are equalities between formal series in powers of ε; this fact will not be pointed out explicitly at each instance. It should also be mentioned that it is possible to envisage an alternative approach where B-series are considered as formal series of terms indexed by $\mathcal{T} \cup \{\emptyset\}$ rather than as formal series of powers of ε; with such an approach —that will not be taken up here— it is not necessary to investigate the convergence of the inner sum in (18).

Given a B-series $B(\delta, y)$ with $\delta_\emptyset = 1$ (i.e. a near-identity B-series), and an arbitrary B-series $B(\eta, y)$, it may be shown (see [15] for the case where the vertices of the trees are not colored) that the composition

$$B(\eta, B(\delta, y)) \tag{19}$$

is a B-series of the form $B(\zeta, y)$, where ζ *only depends on* δ *and* η. More precisely each ζ_u is an explicitly known polynomial in the δ_v's and η_w's, $v, w \in \mathcal{T} \cup \{\emptyset\}$ (see the Appendix for additional details together with an example). We write $\zeta = \delta * \eta$; and thus the non-commutative product $*$ among mappings [8] $\delta, \eta \in \mathbb{C}^{\mathcal{T}\cup\{\emptyset\}}$, ($\delta_\emptyset = 1$) corresponds to the composition of the associated series.

The set of all near-identity B-series is a group with respect to the composition (19) and therefore the subset $\mathcal{G} \subset \mathbb{C}^{\mathcal{T}\cup\{\emptyset\}}$ consisting of the mappings $\delta \in \mathbb{C}^{\mathcal{T}\cup\{\emptyset\}}$ with $\delta_\emptyset = 1$ is a non-commutative group for the product $*$. The group $\mathcal{G}$ is called the *coefficient or Butcher group* [3]. The unit of this group will be denoted by $\mathbb{1}$; of course, $\mathbb{1}_\emptyset = 1$ and $\mathbb{1}_u = 0$ for each $u \in \mathcal{T}$.

To conclude this subsection we shall show how to derive Proposition 2.1 through the use of the product $*$. The manipulations involved will illustrate the techniques to be used later to prove the main results. The aim is to express the exact solution of (1)–(2) as $y(t) = B(\alpha(t), y_0)$. Our first task is to rewrite the function f in (11) itself as a B-series

$$\varepsilon f(y, \theta) = B(\beta(\theta), y), \qquad \varepsilon \sum_{\mathbf{k}\in\mathbb{Z}^d} e^{i\mathbf{k}\cdot\theta} f_{\mathbf{k}}(y) = \sum_{u\in\mathcal{T}} \frac{\varepsilon^{|u|}}{\sigma_u} \beta_u(\theta) \mathcal{F}_u(y),$$

with coefficients $\beta_u(\theta)$ (parameterized by the angular variable $\theta \in \mathbb{T}^d$) defined as follows:

$$\beta_u(\theta) := \begin{cases} e^{i\mathbf{k}\cdot\theta} & \text{if } u = \textcircled{\scriptsize k} \text{ for some } \mathbf{k} \in \mathbb{Z}^d, \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

Recalling that the substitution of y in f corresponds to the product $*$ of the corresponding coefficient maps α and β, the initial value problem (1)–(2) may be rewritten in terms of B-series as:

$$\begin{aligned} \frac{d}{dt} B(\alpha(t), y_0) &= B(\alpha(t) * \beta(t\omega), y_0), \\ B(\alpha(0), y_0) &= B(\mathbb{1}, y_0). \end{aligned}$$

In turn this may be rewritten as an initial-value problem *in the coefficient group,* [9] i.e. a problem whose unknown is a map $\alpha : \mathbb{R} \to \mathcal{G}$:

$$\frac{d}{dt}\alpha(t) = \alpha(t) * \beta(t\omega), \qquad \alpha(0) = \mathbb{1}. \tag{21}$$

[8] If A and B are sets, the notation B^A means the set of all mappings $\phi : A \to B$.

[9] This kind of initial-value problem was first considered in [7].

Note that we have now a *universal* formulation, i.e. one where the choice of f and y_0 in (1)–(2) plays no role. Using the expression for the product $*$ (see the Appendix) we have

$$\forall \mathbf{k} \in \mathbb{Z}^s, \qquad \frac{d}{dt}\alpha_{\circled{\mathbf{k}}}(t) = \beta_{\circled{\mathbf{k}}}(t\omega)$$

and

$$\forall u = [u_1 \cdots u_n]_{\mathbf{k}}, \qquad \frac{d\alpha_u(t)}{dt} = \beta_{\circled{\mathbf{k}}}(t\omega)\alpha_{u_1}(t)\ldots\alpha_{u_n}(t).$$

After taking into account the initial condition $\alpha_u(0) = 0$ for each $u \in \mathcal{T}$, we recover the formulae (16)–(17) found above.

2.4 The transport equation

This subsection contains a number of technical results that play a key role in the proof of the main results.

We begin with a simple auxiliary result which is a consequence of the density in $\mathbb{T}^d$ of the quasi-periodic flow $t \mapsto t\omega$ with a non-resonant vector of frequencies $\omega \in \mathbb{R}^d$:

Lemma 2.4 *Let $w \in \mathbb{C}^{\mathbb{R}\times\mathbb{T}^d}$ be a continuous function that, for each fixed θ, is a polynomial in t. If for all $t \in \mathbb{R}$, $w(t, t\omega) = 0$, then $w(t, \theta) \equiv 0$.*

The following definition will be helpful:

Definition 2.5 *We shall say that a function $w : \mathbb{R} \times \mathbb{T}^d \to \mathbb{C}$ is* polynomial *if there exists a complex polynomial $P \in \mathbb{C}[X_0, X_1, \ldots, X_{2d}]$ in $2d+1$ variables X_0,…,X_{2d}, such that*

$$w(t, \theta) = P(t, e^{i\theta_1}, \ldots, e^{i\theta_d}, e^{-i\theta_1}, \ldots, e^{-i\theta_d}).$$

Furthermore, $\gamma : \mathbb{R} \times \mathbb{T}^d \to \mathcal{G}$ is a polynomial *map, if γ_u is a polynomial function for each $u \in \mathcal{T}$.*

It is easily established by induction in $|u|$, that, for each of the coefficients $\alpha_u(t)$ that appear in the expansion (15), there exists a (uniquely-defined) polynomial function $\gamma_u(t, \theta)$ such that we may write[10]

$$\alpha_u(t) = \gamma_u(t, t\omega). \tag{22}$$

Then $\gamma \in \mathcal{G}^{\mathbb{R}\times\mathbb{T}^d}$ is a polynomial map in the sense of the definition. For instance, for $u = \begin{smallmatrix}\circled{\mathbf{k}}\\ \circled{\mathbf{l}}\end{smallmatrix}$ with $\mathbf{k} = -\mathbf{l} \neq \mathbf{0}$, a simple computation of the iterated integral in the second row of Table 1 shows that

$$\alpha_u(t) = \frac{it}{\mathbf{l}\cdot\omega} + \frac{1 - e^{it\mathbf{l}\cdot\omega}}{(\mathbf{l}\cdot\omega)^2}$$

and therefore

$$\gamma_u(t, \theta) = \frac{it}{\mathbf{l}\cdot\omega} + \frac{1 - e^{i\mathbf{l}\cdot\theta}}{(\mathbf{l}\cdot\omega)^2}.$$

We shall derive a number of properties of the γ_u's.

Proposition 2.6 *If $\gamma \in \mathcal{G}^{\mathbb{R}\times\mathbb{T}^d}$ is defined in (22), then $\gamma(0, \mathbf{0}) = \mathbb{1}$ and*

$$\partial_t\gamma(t, \theta) + \omega \cdot \nabla_\theta\gamma(t, \theta) = \gamma(t, \theta) * \beta(\theta). \tag{23}$$

[10] Recall that ω is assumed throughout to be non-resonant and that the resonant case may be rewritten in non-resonant form by lowering the value of d.

Proof: The chain rule shows that, at each t and $\theta = t\omega$, the equality (23) is valid since $\alpha(t)$ satisfies the differential equation in (21). It is then sufficient to invoke Lemma 2.4 to see that (23) holds at each t and θ. The condition $\gamma(0, \mathbf{0}) = \mathbb{1}$ comes from the initial condition in (21). □

While the transport equation (23) possesses infinitely many solutions that satisfy the condition $\gamma(0, \mathbf{0}) = \mathbb{1}$,[11] we shall establish below that only one of them is a polynomial map.

Lemma 2.7 *Given a polynomial function $w : \mathbb{R} \times \mathbb{T}^d \to \mathbb{C}$, there exists a unique polynomial solution of*

$$\partial_t z(t, \theta) + \omega \cdot \nabla_\theta z(t, \theta) = w(t, \theta), \qquad z(0, \mathbf{0}) = 0. \tag{24}$$

Proof: Consider the Fourier expansions

$$w(t, \theta) = \sum_{\mathbf{k} \in I \subset \mathbb{Z}^d} \hat{w}_{\mathbf{k}}(t)\, e^{i\mathbf{k}\cdot\theta}, \qquad z(t, \theta) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{z}_{\mathbf{k}}(t)\, e^{i\mathbf{k}\cdot\theta},$$

where, by assumption, I is a finite subset of $\mathbb{Z}^d$, and for each $\mathbf{k} \in I$, $\hat{w}_{\mathbf{k}}(t)$ is a polynomial in t. The transport equation in (24) then reads

$$\frac{d}{dt}\hat{z}_{\mathbf{k}}(t) + i\,(\mathbf{k} \cdot \omega)\hat{z}_{\mathbf{k}}(t) = \hat{w}_{\mathbf{k}}(t), \qquad \mathbf{k} \in \mathbb{Z}^d. \tag{25}$$

Clearly, (25) uniquely determines the polynomials $\hat{z}_{\mathbf{k}}(t)$ for $\mathbf{k} \neq \mathbf{0}$. In particular, $\hat{z}_{\mathbf{k}}(t) = 0$ if $\mathbf{k} \in \mathbb{Z}^d \backslash I$. The condition $z(0, \mathbf{0}) = 0$ then becomes

$$\hat{z}_{\mathbf{0}}(0) = -\sum_{\mathbf{k} \in I \backslash \{\mathbf{0}\}} \hat{z}_{\mathbf{k}}(0). \tag{26}$$

For $\mathbf{k} = \mathbf{0}$, the polynomial $\hat{z}_{\mathbf{0}}(t)$ is determined from (25)–(26). □

Proposition 2.8 *There exists a unique polynomial solution $\gamma \in \mathcal{G}^{\mathbb{R}\times\mathbb{T}^d}$ of (23) such that $\gamma(0, \mathbf{0}) = \mathbb{1}$.*

Proof: The equation (23) together with $\gamma(0, \mathbf{0}) = \mathbb{1}$ is equivalent to having, for all $u \in \mathcal{T}$,

$$\partial_t \gamma_u(t, \theta) + \omega \cdot \nabla_\theta \gamma_u(t, \theta) = e^{i\mathbf{k}\cdot\theta}\gamma'_u(t, \theta), \quad \gamma_u(0, \mathbf{0}) = 0, \tag{27}$$

where $\gamma'_u(t, \theta) = 1$ if $|u| = 1$, and

$$\gamma'_u(t, \theta) = \prod_{\nu=1}^{n} \gamma_{u_\nu}(t, \theta)$$

for u in (14). It is then enough to prove that for each $u \in \mathcal{T}$, there exists a unique a polynomial function $\gamma_{\,u}(t, \theta)$ satisfying (27). This can readily be proved by induction on $|u|$: For $|u| = 1$, application of Lemma 2.7 immediately gives the required result. For $|u| > 1$, the induction hypothesis shows that the function $e^{\,i\mathbf{k}\cdot\theta}\gamma'_u(t, \theta)$ is polynomial, so that applying Lemma 2.7 once more leads to the required result. □

The last result in this subsection is essential in later developments:

Proposition 2.9 *The map γ defined in (22) (or in Proposition 2.8) satisfies the identities*

$$\forall t, t' \in \mathbb{R}, \qquad \gamma(t', \mathbf{0}) * \gamma(t, \mathbf{0}) = \gamma(t' + t, \mathbf{0}) \tag{28}$$

and

$$\forall t \in \mathbb{R},\ \forall \theta \in \mathbb{T}^d, \qquad \gamma(t, \mathbf{0}) * \gamma(0, \theta) = \gamma(t, \theta). \tag{29}$$

[11] The method of characteristics shows that the solutions of (23) with $\gamma(0, \mathbf{0}) = \mathbb{1}$ are defined by

$$\gamma(t, \theta) = \chi(\theta - t\omega) + \int_0^t \gamma(s, \theta + (s - t)\omega) * \beta(\theta + (s - t)\omega)\, ds.$$

The initial-value function $\chi : \mathbb{T}^d \to \mathcal{G}$ has to vanish at $\theta = \mathbf{0}$, but it is otherwise arbitrary.

Proof: Both identities may be combined into one:

$$\forall t, t' \in \mathbb{R},\ \forall \theta \in \mathbb{T}^d, \qquad \gamma(t', \mathbf{0}) * \gamma(t, \theta) = \gamma(t' + t, \theta).$$

Consider, for fixed t', the map $\tilde{\gamma} \in \mathcal{G}^{\mathbb{R}\times\mathbb{T}^d}$ defined as $\tilde{\gamma}(t, \theta) = \gamma(t', \mathbf{0})^{-1} * \gamma(t' + t, \theta)$. By the linearity of $*$ with respect to the right factor (see the Appendix), $\tilde{\gamma}(t, \theta)$ is a solution of (23); it is furthermore polynomial, and trivially satisfies $\tilde{\gamma}(0, \mathbf{0}) = \mathbb{1}$. By Proposition 2.8, $\tilde{\gamma}(t, \theta) \equiv \gamma(t, \theta)$. □

2.5 Quasi-stroboscopic averaging

We are now ready to describe the technique of quasi-stroboscopic averaging. The identity (28) together with $\gamma(0, \mathbf{0}) = \mathbb{1}$ shows that the mappings

$$\bar{\alpha}(t) := \gamma(t, \mathbf{0}), \tag{30}$$

parameterized by $t \in \mathbb{R}$, form a one-parameter subgroup of the coefficient group $\mathcal{G}$. Therefore $\bar{\alpha}(t)$ is the solution of the *autonomous* initial-value problem (cf. (21))

$$\frac{d}{dt}\bar{\alpha}(t) = \bar{\alpha}(t) * \bar{\beta}, \qquad \bar{\alpha}(0) = \mathbb{1}, \tag{31}$$

with

$$\bar{\beta} := \left.\frac{d}{dt}\bar{\alpha}(t)\right|_{t=0}. \tag{32}$$

(For completeness, we include here a proof, but the result is of course standard in the theory of differential equations:

$$\frac{d}{dt'}\bar{\alpha}(t') = \left.\frac{d}{dt}\bar{\alpha}(t' + t)\right|_{t=0} = \left.\frac{d}{dt}\bar{\alpha}(t') * \bar{\alpha}(t)\right|_{t=0} = \bar{\alpha}(t') * \bar{\beta};$$

when differentiating the product $*$ we have taken into account once more that it is linear with respect to the right factor, see the Appendix.) According to (29) the solution $\alpha(t) = \gamma(t, t\omega)$ of the oscillatory problem (21) may be represented as

$$\alpha(t) = \bar{\alpha}(t) * \kappa(t\omega)$$

in terms of the solution of the autonomous problem (31) and the map $\kappa \in \mathcal{G}^{\mathbb{T}^d}$ defined by

$$\kappa(\theta) := \gamma(0, \theta). \tag{33}$$

After plugging the maps $\alpha(t)$, $\bar{\alpha}(t)$, $\bar{\beta}$, and $\kappa(\theta)$ into the corresponding B-series, we conclude:

Theorem 2.10 *The solution y of (1)–(2) may be written as*

$$y(t) = U(Y(t), t\omega) \tag{34}$$

where U is the change of variables (parameterized by $\theta \in \mathbb{T}^s d$)

$$U(Y, \theta) := B(\kappa(\theta), Y) := Y + \sum_{u\in\mathcal{T}} \frac{\varepsilon^{|u|}}{\sigma_u}\kappa_u(\theta)\, \mathcal{F}_u(Y) \tag{35}$$

and $Y(t)$ is the solution of the (averaged) autonomous initial value problem

$$\frac{d}{dt}Y = \varepsilon F(Y), \qquad Y(0) = y_0, \tag{36}$$

with

$$\varepsilon F(Y) := B(\bar{\beta}, Y) := \sum_{u\in\mathcal{T}} \frac{\varepsilon^{|u|}}{\sigma_u}\bar{\beta}_u\, \mathcal{F}_u(Y). \tag{37}$$

$$
\begin{array}{ccccc}
\frac{d}{dt}y &=& \varepsilon f(y,t\omega) & & \\
f(y,\theta) &=& B(\beta(\theta),y) & & \\
\downarrow \text{(16)–(17)} & & & & \\
y(t) &=& B(\alpha(t),y_0) & & \\
\alpha(t) &=& \gamma(t,t\omega) & \rightarrow & \\
\downarrow & & & & \\
y(t) &=& U(Y(t),t\omega) & & \\
U(Y,\theta) &=& B(\kappa(\theta),Y) & \rightarrow & \\
\kappa(\theta) &=& \gamma(0,\theta) & &
\end{array}
\qquad
\begin{array}{ccc}
\frac{d}{dt}Y &=& \varepsilon F(Y) \\
F(Y) &=& B(\bar\beta,Y) \\
\uparrow \text{(32)} & & \\
Y(t) &=& B(\bar\alpha(t),y_0) \\
\bar\alpha(t) &=& \gamma(t,\mathbf{0}) \\
& & \\
\widehat{Y}(t) &=& \widehat{U}(\widehat{Y}(t),t\omega) \\
\widehat{U}(\widehat{Y},\theta) &=& B(\widehat{\kappa}(\theta),\widehat{Y}) \\
\widehat{\kappa}(\theta) &=& \lambda * \kappa(\theta)
\end{array}
\qquad
\begin{array}{ccc}
\frac{d}{dt}\widehat{Y} &=& \varepsilon \widehat{F}(\widehat{Y}) \\
\widehat{F}(\widehat{Y}) &=& B(\widehat{\bar\beta},\widehat{Y}) \\
\uparrow \text{(43)} & & \\
\widehat{Y}(t) &=& B(\widehat{\bar\alpha}(t),\widehat{Y}(0)) \\
\widehat{\bar\alpha}(t) &=& \lambda * \bar\alpha(t) * \lambda^{-1} \\
\uparrow & & \\
Y &=& V(\widehat{Y}) \\
Y &=& B(\lambda,\widehat{Y})
\end{array}
\leftarrow
$$

Table 2: Overview of Section 2. The table consists of three blocks of rows. The first row of the top block contains the original oscillatory system with solution y, the quasi-stroboscopically averaged system with solution Y and a general B-series averaged system with solution $\widehat{Y}$ (as in Section 2.6). The second row of the top block provides the B-series for the corresponding vector fields f, F, $\widehat{F}$. The middle block refers to the solutions y, Y, $\widehat{Y}$ and the bottom block to the changes of variables that relate these solutions. The computation starts from the given $\beta(\theta)$ in the top left corner; the arrows and equation numbers show how to reach the coefficients $\bar\beta$, $\kappa(\theta)$ of the stroboscopically averaged system and the associated change of variables. For general B-series averaged systems (Section 2.6) the change V in the bottom right corner may be chosen arbitrarily; again the arrows and equation numbers show how to reach the coefficients $\widehat{\bar\beta}$, $\widehat{\kappa}(\theta)$

Here $\bar\beta$ and $\kappa(\theta)$ are as defined in (32) and (33).

Furthermore the averaged solution Y possesses an expansion

$$
Y(t) = B(\bar\alpha(t),y_0) = y_0 + \sum_{u\in\mathcal{T}} \frac{\varepsilon^{|u|}}{\sigma_u}\bar\alpha_u(t)\,\mathcal{F}_u(y_0) \tag{38}
$$

with $\bar\alpha(t)$ defined in (30).

Table 2 shows the path we have followed to define and compute the coefficients $\bar\beta_u$ of the averaged system and $\kappa_u(\theta)$ of the change of variables; these coefficients are *universal* in the sense mentioned before. For the averaged system we find, after computing β_u for $|u| = 1, 2$:

$$
\frac{d}{dt}Y \;=\; \varepsilon f_{\mathbf{0}} + \varepsilon^2 \sum_{\mathbf{k}\neq\mathbf{0}} \frac{i}{\mathbf{k}\cdot\omega}\big(f'_{\mathbf{0}}(Y)f_{\mathbf{k}}(Y) - f'_{\mathbf{k}}(Y)f_{\mathbf{0}}(Y) + f'_{\mathbf{k}}(Y)f_{-\mathbf{k}}(Y)\big) + \mathcal{O}(\varepsilon^3).
$$

This is the extension to the multi-frequency case of formula (28) in [10]. The ε^3 terms will be displayed in formula (64). Note that with our approach the sets of coefficients $\bar\beta_u$ and $\kappa_u(\theta)$ may be found independently of each other, while in (7)–(10) the u_j's and F_j's are coupled.

The oscillatory and averaged solutions $y(t)$ and $Y(t)$ coincide at the initial time $t = 0$, because $\kappa(\mathbf{0}) = \gamma(0,\mathbf{0}) = \mathbb{1}$. In the single frequency case, $d = 1$, the equality (34) implies that $y(t)$ and $Y(t)$ also coincide at the stroboscopic times $t = n(2\pi/\omega)$, n integer; this is the situation considered in [10]. However when $d > 1$ the angle $t\omega$ never returns to its initial value $\mathbf{0} \in \mathbb{T}^d$ and the stroboscopic effect is not present. For $d > 1$ we shall refer to the averaging technique described in the theorem as *quasi-stroboscopic averaging*, because, for arbitrarily small $\mu > 0$, there exist quasi-periods $T_\mu > 0$ such that $U(Y, nT_\mu\omega) = Y + \mathcal{O}(n\mu)$ and thus $y(nT_\mu) = Y(nT_\mu) + \mathcal{O}(n\mu)$.

2.6 General high-order averaging

The quasi-stroboscopic averaged system (36)–(37) is by no means the only averaged version of the oscillatory system (1); it is always possible to transform (36)–(37) by changing variables in $\mathbb{R}^D$. The proof of our next result

is standard and will not be given. While in Theorem 2.10 we spelled out the B-series that appeared, we shall not do so hereafter.

Proposition 2.11 *(i) Under a smooth invertible change of variables*

$$Y = V(\widehat{Y}),$$

the initial-problem (36)–(37) of quasi-stroboscopic averaging is transformed into

$$\frac{d}{dt}\widehat{Y} = \varepsilon \widehat{F}(\widehat{Y}), \qquad V(\widehat{Y}(0)) = y_0, \tag{39}$$

with $\widehat{F}$ given by the pull-back

$$\widehat{F}(\widehat{Y}) := \left(\frac{\partial V(\widehat{Y})}{\partial \widehat{Y}}\right)^{-1} F(V(\widehat{Y})). \tag{40}$$

If y solves the oscillatory initial value problem (1)–(2), then (cf. (34))

$$y(t) = \widehat{U}(\widehat{Y}(t), t\omega), \tag{41}$$

where $\widehat{Y}$ is the solution of (39) and $\widehat{U}$ is the change of variables (parameterized by $\theta \in \mathbb{T}^s$)

$$\widehat{U}(\widehat{Y}, \theta) := U(V(\widehat{Y}), \theta).$$

(Here U is as in (35).)

(ii) In the particular case where V is given by a B-series, i.e.

$$V(\widehat{Y}) = B(\lambda, \widehat{Y})$$

with $\lambda \in \mathcal{G}$, the change $\widehat{U}$ is a near-identity B-series (parameterized by $\theta \in \mathbb{T}^s$)(cf. (35)):

$$\widehat{U}(\widehat{Y}, \theta) = B(\widehat{\kappa}(\theta), \widehat{Y}), \qquad \widehat{\kappa}(\theta) := \lambda * \kappa(\theta), \tag{42}$$

the solution $\widehat{Y}(t)$ satisfies (cf. (38))

$$\widehat{Y}(t) = B(\widehat{\bar{\alpha}}(t), \widehat{Y}(0)), \qquad \widehat{\bar{\alpha}}(t) := \lambda * \bar{\alpha}(t) * \lambda^{-1},$$

and (cf. (32))

$$\varepsilon \widehat{F}(\widehat{Y}) = B(\widehat{\bar{\beta}}, \widehat{Y}), \qquad \widehat{\bar{\beta}} := \frac{d}{dt}\widehat{\bar{\alpha}}(t)\bigg|_{t=0}. \tag{43}$$

Table 2 describes the computation of all required B-series starting from the given original system (1) and the chosen λ.

Remark 2.12 In (7)–(10), the change of variables (41) is determined to ensure

$$\frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} \widehat{U}(\widehat{Y}, \theta)\, d\theta = \widehat{Y}.$$

By taking angular averages on both sides of the equality (42) this requirement leads to the condition $\mathbb{1} = \lambda * \chi$ with

$$\chi = \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} \kappa(\theta)\, d\theta. \tag{44}$$

Thus (7)–(10) is the instance of Proposition 2.11 (ii) that corresponds to the choice $\lambda = \chi^{-1}$.

Remark 2.13 In part (ii) the coefficients $\widehat{\bar{\beta}}_u$ are computed from the coefficients $\widehat{\bar{\alpha}}_u(t)$. It is also possible to express the $\widehat{\bar{\beta}}_u$ in terms of the $\bar{\beta}_u$ by changing variables in (39) (refer to Table 2), the recipe is

$$\widehat{\bar{\beta}} = \lambda * \bar{\beta} * \lambda^{-1}. \tag{45}$$

The product in the right-hand side cannot be interpreted in terms of compositions as in (19) because $\bar{\beta} \notin \mathcal{G}$. The derivation of (45) and the interpretation of the right-hand side require to think of λ and $\bar{\beta}$ as complex-valued applications defined in a set of forests (rather than rooted trees), as in [8] or [9]. [12]

It is remarkable that *all* averaged systems may be obtained by choosing appropriately V (or λ) in Proposition 2.11. We consider first the case where the change of variables $y \mapsto \widehat{Y}$ is assumed to be given by a B-series:

Theorem 2.14 *Suppose that the change of variables (41) with*

$$\widehat{U}(\widehat{Y}, \theta) = B(\widehat{\kappa}(\theta), \widehat{Y}), \qquad \widehat{\kappa} \in \mathcal{G}^{\mathbb{T}^d},$$

transforms the oscillatory differential system (1) into an (averaged) autonomous differential system. Then the averaged system and the change $\widehat{U}$ are those obtained by applying Proposition 2.11 (ii) with $\lambda = \widehat{\kappa}(\mathbf{0})$.

Proof: The averaged solution satisfies $\widehat{Y}(t) = B(\widehat{\bar{\alpha}}(t), \widehat{Y}(0))$ for a suitable polynomial map $t \mapsto \widehat{\bar{\alpha}}(t) \in \mathcal{G}$ with $\widehat{\bar{\alpha}}(0) = \mathbb{1}$, while $y(t) = B(\gamma(t, t\omega), y(0))$. Then the hypothesis of the theorem may be expressed as:

$$B(\widehat{\bar{\alpha}}(t) * \widehat{\kappa}(t\omega), y_0) = B(\widehat{\kappa}(\mathbf{0}) * \gamma(t, t\omega), y_0)$$

for an arbitrary initial value y_0. Lemma 2.4 yields:

$$B(\widehat{\bar{\alpha}}(t) * \widehat{\kappa}(\theta), y_0) = B(\widehat{\kappa}(\mathbf{0}) * \gamma(t, \theta), y_0),$$

and, in particular, for $\theta = \mathbf{0}$,

$$B(\widehat{\bar{\alpha}}(t) * \widehat{\kappa}(\mathbf{0}), y_0) = B(\widehat{\kappa}(\mathbf{0}) * \bar{\alpha}(t), y_0),$$

and thus $\hat{U}(\widehat{Y}, \theta) = U(V(\widehat{Y}), \theta)$ with $V(\widehat{Y}) = B(\widehat{\kappa}(\mathbf{0}), \widehat{Y})$. □

We study next the situation when the change $\widehat{U}$ is not necessarily given by a B-series.

Theorem 2.15 *Assume that the change of variables (41) with*

$$\widehat{U}(\widehat{Y}, \theta) = \widehat{Y} + \varepsilon \widehat{u}_1(\widehat{Y}, \theta) + \varepsilon^2 \widehat{u}_2(\widehat{Y}, \theta) + \cdots \tag{46}$$

transforms the oscillatory differential system (1) into an (averaged) autonomous differential system $(d/dt)\,\widehat{Y} = \varepsilon \widehat{F}(\widehat{Y})$. Then there is a near-identity change of variables $Y = V(\widehat{Y})$ such that (39) and (40) hold, i.e. the averaged problem is obtained by changing variables in the quasi-stroboscopic averaged problem (36)–(37). Furthermore $V(\cdot) = \widehat{U}(\cdot, \mathbf{0})$.

Proof: We carry out the proof first for the case where $\widehat{U}(\widehat{Y}, \mathbf{0}) \equiv \widehat{Y}$, i.e. $\widehat{Y}$ and y coincide at $t = 0$. Let $\widehat{Y} = \Phi_t(y_0)$ be the solution flow of the averaged system; by expanding Φ_t in powers of ε and substituting into (46) we obtain

$$\widehat{U}(\widehat{Y}(t), \theta) = y_0 + \varepsilon G_1(t, \theta, y_0) + \varepsilon^2 G_2(t, \theta, y_0) + \cdots$$

where each G_j is polynomial for each fixed value of the variable y_0. By hypothesis $\widehat{U}(\widehat{Y}(t), t\omega)$ must coincide with $y(t) = B(\gamma(t, t\omega), y_0)$ and thus, for each j,

$$G_j(t, \theta, y_0) = \sum_{|u|=j} \frac{1}{\sigma_u} \gamma_u(t, \theta)\, \mathcal{F}_u(y_0)$$

[12] The use of forests is essential in establishing the connection between formal series and Hopf algebras, see e.g. [4], [19]. That connection clarifies greatly the computation of the product $*$.

whenever $\theta = t\omega$; application of Lemma 2.4 shows that the equality holds in fact for all t and θ. Thus

$$\widehat{U}(\widehat{Y}(t), \theta) = B(\gamma(t, \theta), y_0)$$

and, in particular, at $t = 0$,

$$\widehat{U}(y_0, \theta) = B(\kappa(\theta), y_0).$$

Here we recognize that the change $\widehat{U}$ is in fact the one (35) that transforms the original system (1) into the quasi-stroboscopically averaged (36)–(37) and the proof is complete.

When $\widehat{U}(\cdot, \mathbf{0}) =: V(\cdot)$ is not the identity map the proof is similar, but an initial change of variables $V(\widehat{Y}) = Z$ is required. □

Remark 2.16 The preceding results show that $V(\cdot) = \widehat{U}(\cdot, \mathbf{0})$ is the only 'free parameter' in the whole procedure. All possible averaged systems are obtained by performing the (time-independent) change of variables $Y = V(\widehat{Y})$ in the quasi-stroboscopic system (36)–(37). If V happens to be given by a B-series, then the averaged system resulting from the change of variables has a B-series right-hand side and the relation (41) will hold for a function $\widehat{U}$ expressible as a B-series.

2.7 Changing the initial time

If the initial condition that supplements the system (1) is prescribed at an initial time $t_0 \neq 0$,

$$y(t_0) = y_0, \tag{47}$$

it may be of interest to look for a change of variables $y(t) = \widetilde{U}(\widetilde{Y}, t\omega)$ that satisfies

$$\widetilde{U}(\widetilde{Y}, t_0\omega) \equiv \widetilde{Y} \tag{48}$$

(instead of $U(Y, \mathbf{0}) \equiv Y$) so as to ensure that (47) leads to an initial condition $\widetilde{Y}(t_0) = y_0$.

If $\widetilde{U}$ is sought as a B-series, $\widetilde{U}(\widetilde{Y}, \theta) = B(\widetilde{\kappa}(\theta), \widetilde{Y})$, the condition (48) becomes $\widetilde{\kappa}(t_0\omega) = \mathbb{1}$ or, in the framework of Proposition 2.11 (see (42)), $\lambda * \kappa(t_0\omega) = \mathbb{1}$. Therefore the required change of variables has coefficients

$$\widetilde{\kappa}(\theta) = \kappa(t_0\omega)^{-1} * \kappa(\theta) \tag{49}$$

and, if we define

$$\widetilde{\bar{\alpha}}(t) := \kappa(t_0\omega)^{-1} * \bar{\alpha}(t) * \kappa(t_0\omega) \tag{50}$$

then $\widetilde{Y}(t) = B(\widetilde{\bar{\alpha}}(t), \widetilde{Y}(0))$ or, since the averaged problem is autonomous

$$\widetilde{Y}(t) = B(\widetilde{\bar{\alpha}}(t - t_0), y_0).$$

The coefficients $\widetilde{\bar{\beta}}_u$ for the right-hand side $\widetilde{F}$ of the averaged equation may be found either by differentiation of $\widetilde{\bar{\alpha}}_u(t)$ or as in Remark 2.13.

In the proposition below, we shall derive alternative expressions for $\widetilde{\kappa}(\theta)$, $\widetilde{\bar{\alpha}}(t)$ and $\widetilde{\bar{\beta}}$. Some notation and a lemma are required. The *total multi-index* $\mathcal{I}(u) \in \mathbb{Z}^d$ of a tree $u \in \mathcal{T}$ is defined as the sum of all the multi-indices that label its vertices. More precisely

1. For each tree $\tau_{\mathbf{k}}$ of order 1, $\mathcal{I}(u) := \mathbf{k}$.
2. For u in (14),

$$\mathcal{I}(u) := \mathbf{k} + \mathcal{I}(u_1) + \cdots + \mathcal{I}(u_n),$$

If $\delta \in \mathbb{C}^{\mathcal{T}\cup\{\emptyset\}}$ and $\theta \in \mathbb{T}^d$, we define $\delta^\theta \in \mathbb{C}^{\mathcal{T}\cup\{\emptyset\}}$ as follows. For each $u \in \mathbb{T}$,

$$\delta^\theta_u = e^{i\mathcal{I}(u)\cdot\theta}\delta_u, \tag{51}$$

and $\delta^\theta_\emptyset = \delta_\emptyset$. It follows (see the Appendix) that, for $\delta, \eta \in \mathbb{C}^{\mathcal{T}\cup\{\emptyset\}}$, with $\delta_\emptyset = 1$, and $\theta \in \mathbb{T}^d$,

$$(\delta * \eta)^\theta = \delta^\theta * \eta^\theta. \tag{52}$$

Lemma 2.17 *For each $t \in \mathbb{R}$ and each $\theta, \theta_0 \in \mathbb{T}^d$,*

$$\gamma(0,\theta_0)^{-1} * \gamma(t, \theta+\theta_0) = \gamma^{\theta_0}(t,\theta). \tag{53}$$

Proof: By definition of $\beta(\theta)$ (see (20)), $\beta(\theta+\theta_0) = \beta^{\theta_0}(\theta)$, and then (52) and (23) imply that $\tilde\gamma(t,\theta) := \gamma^{\theta_0}(t,\theta)$ solves the transport equation

$$\partial_t\tilde\gamma(t,\theta) + \omega\cdot\nabla_\theta\tilde\gamma(t,\theta) = \tilde\gamma(t,\theta) * \beta(\theta+\theta_0).$$

Furthermore $\tilde\gamma(0,\mathbf{0}) = \mathbb{1}$. On the other hand, $\gamma(0,\theta_0)^{-1} * \gamma(t,\theta+\theta_0)$ satisfies the same transport equation and also takes the value $\mathbb{1}$ at $t=0$, $\theta = \mathbf{0}$. The proof is concluded by applying Lemma 2.7 as in Proposition 2.8. □

Proposition 2.18 *With the notation above,*

$$\widetilde{\kappa}(\theta) = \kappa^{t_0\omega}(\theta - t_0\omega), \qquad \widetilde{\bar\alpha}(t) = \bar\alpha^{t_0\omega}(t), \qquad \widetilde{\bar\beta} = \frac{d}{dt}\widetilde{\bar\alpha}(t)\Big|_{t=0}.$$

Proof: The formula for $\widetilde{\bar\beta}$ is obviously a consequence of the formula for $\widetilde{\bar\alpha}(t)$. From (49)–(50) and the definitions of $\bar\alpha$ and κ in (30) and (33), we see that it remains to prove that

$$\gamma(0,t_0\omega)^{-1} * \gamma(0,\theta) = \gamma^{t_0\omega}(0,\theta-t_0\omega), \qquad \gamma(0,t_0\omega)^{-1} * \gamma(t,\mathbf{0}) * \gamma(0,t_0\omega) = \gamma^{t_0\omega}(t,\mathbf{0}).$$

The identity (29) makes it possible to rewrite the last equality in the simpler form

$$\gamma(0,t_0\omega)^{-1} * \gamma(t,t_0\omega) = \gamma^{t_0\omega}(t,\mathbf{0}),$$

and, accordingly, the proposition is a consequence of the preceding lemma. □

3 Geometric properties of quasi-stroboscopic averaging

The coefficient group $(\mathcal{G},*)$ may be regarded (in a formal way) as an infinite-dimensional Lie group. The corresponding Lie algebra is given by

$$\mathfrak{g} = \{\beta \in \mathbb{C}^{\mathcal{T}\cup\{\emptyset\}} : \beta_\emptyset = 0\},$$

because $\mathfrak{g}$ is the set of all possible values of the velocity $(d\alpha/dt)(0)$ of the curves $t \mapsto \alpha(t) \in \mathcal{G}$ that satisfy $\alpha(0) = \mathbb{1}$.[13] Following Murua [19] we may consider the subgroup $\widehat{\mathcal{G}} \subset \mathcal{G}$ of those $\delta \in \mathcal{G} \subset \mathbb{C}^{\mathcal{T}\cup\{\emptyset\}}$ such that for all $u, v, w \in \mathcal{T}$

$$\delta_{u\circ v} + \delta_{v\circ u} = \delta_u\delta_v, \tag{54}$$

$$\delta_{(u\circ v)\circ w} + \delta_{(v\circ u)\circ w} + \delta_{(w\circ u)\circ v} = \delta_u\delta_v\delta_w, \tag{55}$$

where $\circ$ denotes the Butcher product in $\mathcal{T}$, i.e. $u\circ v$ is the tree obtained by grafting the root of v into the root of u (for u in (14), $u\circ v = [u_1,\dots,u_n,v]_{\mathbf{k}}$). The geometric significance of $\widehat{\mathcal{G}}$ is that for $\delta \in \widehat{\mathcal{G}}$ the processes of forming

[13] In fact (32) and (43) provide examples of the computation of an element in the Lie algebra —infinitesimal generator— in terms of the corresponding one-parameter semigroup.

B-series and changing variables in the vector field commute (see Proposition 3.1 below). In particular, for the true solution y, it is true that for each $t \in \mathbb{R}$, $\alpha(t) \in \widehat{\mathcal{G}}$.

Given $\delta \in \widehat{\mathcal{G}}$, the relations (54)–(55) may be used [19] to group together terms in the B-series $B(\delta, y)$ and obtain the more compact form:

$$B(\delta, y) = y + \sum_{r=1}^{\infty} \sum_{\mathbf{k}_1, \ldots, \mathbf{k}_r \in \mathbb{Z}^d} \varepsilon^r \delta_{\mathbf{k}_1 \cdots \mathbf{k}_r} \, f_{\mathbf{k}_1 \cdots \mathbf{k}_r}(y); \tag{56}$$

here the notation is as follows:

- $\delta_{\mathbf{k}_1 \cdots \mathbf{k}_r} = \delta_{u_{\mathbf{k}_1 \cdots \mathbf{k}_r}}$, where the ('tall' or 'un-ramified') trees $u_{\mathbf{k}_1 \cdots \mathbf{k}_r} \in \mathcal{T}$ are recursively defined by

$$u_{\mathbf{k}} := \textcircled{\scriptsize \mathbf{k}}, \qquad u_{\mathbf{k}_1 \cdots \mathbf{k}_r} := [u_{\mathbf{k}_1 \cdots \mathbf{k}_{r-1}}]_{\mathbf{k}_r}.$$

- $f_{\mathbf{k}_1 \cdots \mathbf{k}_r}(y) := \partial_y f_{\mathbf{k}_2 \cdots \mathbf{k}_r}(y) f_{\mathbf{k}_1}(y)$.

In (56) there is only a term per *word* $\mathbf{k}_1 \cdots \mathbf{k}_r$ over the alphabet $\mathbb{Z}^d$; in (18) there is a term per rooted tree with vertices labelled by the same alphabet. Nevertheless, for $\delta \in \widehat{\mathcal{G}}$, the constraints (54)–(55) still impose some relations among the values of the coefficients $\delta_{u_{\mathbf{k}_1 \cdots \mathbf{k}_r}}$ so that not all of them may be chosen arbitrarily (for instance $2\delta_{\mathbf{kk}} = \delta_{\mathbf{k}}^2$ for each $\mathbf{k} \in \mathbb{Z}^d$).

The Lie algebra $\widehat{\mathfrak{g}}$ of $\widehat{\mathcal{G}}$ is of course a subalgebra of $\mathfrak{g}$; for $\beta \in \widehat{\mathfrak{g}}$ the B-series $B(\beta, y)$ may be rewritten [19] as

$$B(\beta, y) = \sum_{r=1}^{\infty} \sum_{\mathbf{k}_1, \ldots, \mathbf{k}_r \in \mathbb{Z}^d} \varepsilon^r \beta_{\mathbf{k}_1 \cdots \mathbf{k}_r} \, f_{\mathbf{k}_1 \cdots \mathbf{k}_r}(y); \tag{57}$$

where

$$\beta_{\mathbf{k}_1 \cdots \mathbf{k}_r} := \frac{d}{dt} \alpha_{\mathbf{k}_1 \cdots \mathbf{k}_r}(t) \Big|_{t=0} \tag{58}$$

and $t \mapsto \alpha(t) \in \widehat{\mathcal{G}}$ is a smooth curve with $\alpha(0) = \mathbb{1}$. Not all the values $\beta_{\mathbf{k}_1 \cdots \mathbf{k}_r}$ are independent, because of the constraints on the coefficients $\alpha_{\mathbf{k}_1 \cdots \mathbf{k}_r}(t)$ mentioned above (for instance $\beta_{\mathbf{kk}} = 0$ for each $\mathbf{k} \in \mathbb{Z}^d$, since $(d/dt)\alpha_{\mathbf{kk}} = \alpha_{\mathbf{k}}(t)(d/dt)\alpha_{\mathbf{k}}(t)$ and $\alpha_{\mathbf{k}}(0) = 0$). The Dynkin–Specht–Wever theorem [16], implies that (57) may be rearranged to read

$$B(\beta, y) = \sum_{r=1}^{\infty} \sum_{\mathbf{k}_1, \ldots, \mathbf{k}_r \in Z^d} \frac{\varepsilon^r}{r} \, \beta_{\mathbf{k}_1 \cdots \mathbf{k}_r} \, [[\cdots [[f_{\mathbf{k}_1}, f_{\mathbf{k}_2}], f_{\mathbf{k}_3}] \cdots], f_{\mathbf{k}_r}](y), \tag{59}$$

where, for each pair of vector fields, f, g,

$$[f, g](y) = \left(\frac{\partial}{\partial y} g(y) \right) f(y) - \left(\frac{\partial}{\partial y} f(y) \right) g(y)$$

denotes the corresponding commutator (Lie bracket). It should be emphasized that in the representation (59) not all the iterated commutators are linearly independent in view of the skew-symmetry of $[\cdot, \cdot]$ and of the Jacobi identity (for instance $[f_{\mathbf{k}_1}, f_{\mathbf{k}_2}] = -[f_{\mathbf{k}_2}, f_{\mathbf{k}_1}]$ for all $\mathbf{k}_1, \mathbf{k}_2 \in \mathbb{Z}^d$).[14]

Since commutators are geometric objects (i.e. the operations forming commutators and changing variables in a vector field commute), we obtain from (59):

[14] This point is taken up in the next section.

Proposition 3.1 *Consider a change of co-ordinates $y = C(x)$, so that the vector fields $f_{\mathbf{k}}$, $\mathbf{k} \in \mathbb{Z}^d$, in (11) become in the new variables*

$$\tilde{f}_{\mathbf{k}}(x) = \left(\frac{\partial}{\partial x} C(x)\right)^{-1} f_{\mathbf{k}}(C(x)).$$

For each $\delta \in \mathbb{C}^{\mathcal{T} \cup \{\emptyset\}}$, denote by $\tilde{B}(\delta, x)$ the B-series with elementary differentials $\tilde{\mathcal{F}}_u(x)$ built from the vector fields $\tilde{f}_{\mathbf{k}}(x)$ in lieu of the elementary differentials built from the $f_{\mathbf{k}}(y)$.

- *If $\beta \in \widehat{\mathfrak{g}}$, then*

$$\tilde{B}(\beta, x) = \left(\frac{\partial}{\partial x} C(x)\right)^{-1} B(\beta, C(x)),$$

- *If $\delta \in \widehat{\mathcal{G}}$, then*

$$C(\tilde{B}(\delta, x)) = B(\delta, C(x)).$$

If $\delta \in \mathcal{G} \backslash \widehat{\mathcal{G}}$, the commutation in the proposition only holds if $y = C(x)$ is an affine map.

After these preparations, we may state our next theorem, which implies that quasi-stroboscopic averaging is geometric in nature: the quasi-stroboscopic averaging procedure and an arbitrary smooth change of variables in $\mathbb{R}^D$ commute with each other, so that quasi-stroboscopic averaging makes sense (intrinsically) for non-autonomous quasi-periodic smooth ODEs defined on smooth manifolds. The cases of divergence-free and canonical vector fields are singled out in the statement of the theorem; similar results hold of course for any other Lie algebra of vector fields. Another application is to the case of dynamical systems with symmetry, where averaging may be used to construct normal forms for bifurcations.

Theorem 3.2 *(i) The autonomous system in (36)–(37) corresponding to quasi-stroboscopic averaging can be written in the form*

$$\frac{d}{dt} Y = \sum_{r=1}^{\infty} \sum_{\mathbf{k}_1, \ldots, \mathbf{k}_r \in Z^d} \frac{\varepsilon^r}{r} \, \bar{\beta}_{\mathbf{k}_1 \cdots \mathbf{k}_r} \, [[\cdots [[f_{\mathbf{k}_1}, f_{\mathbf{k}_2}], f_{\mathbf{k}_3}] \cdots], f_{\mathbf{k}_r}](Y), \tag{60}$$

where the coefficients $\bar{\beta}_{\mathbf{k}_1 \cdots \mathbf{k}_r}$ are defined from the $\bar{\alpha}_{\mathbf{k}_1 \cdots \mathbf{k}_r}(t) := \bar{\alpha}_{u_{\mathbf{k}_1 \cdots \mathbf{k}_r}}(t)$ by differentiation as in (58).

(ii) Consider the case where the vector field (11) is divergence-free and therefore the solution flow of (1) is volume-preserving. In this case the averaged system (60) is also divergence-free and therefore its solution flow is volume-preserving. In addition, the change of variables (35), i.e. $U(\cdot, \theta) = B(\kappa(\theta), \cdot)$, is volume-preserving for each $\theta \in \mathbb{T}^d$.

(iii) Consider the case where D is even and (11) is the canonical Hamiltonian vector field associated with the Hamiltonian function $H(y, \theta)$ with Fourier expansion

$$H(y, \theta) = \sum_{\mathbf{k} \in \mathbb{Z}^d} e^{i \mathbf{k} \cdot \theta} H_{\mathbf{k}}(y)$$

(so that, for each $\mathbf{k} \in \mathbb{Z}^d$, $f_{\mathbf{k}}(y)$ is the canonical Hamiltonian vector field associated with $H_{\mathbf{k}}$). In this case the averaged system (60) is also canonical, with Hamiltonian function $\varepsilon \bar{H}$ defined in terms of canonical Poisson brackets[15] $\{\cdot, \cdot\}$ of the $H_{\mathbf{k}}$:

$$\varepsilon \bar{H} = \sum_{r=1}^{\infty} \sum_{\mathbf{k}_1, \ldots, \mathbf{k}_r \in Z^d} \frac{\varepsilon^r}{r} \, \bar{\beta}_{\mathbf{k}_1 \cdots \mathbf{k}_r} \, \{\{\cdots \{\{H_{\mathbf{k}_1}, H_{\mathbf{k}_2}\}, H_{\mathbf{k}_3}\} \cdots\}, H_{\mathbf{k}_r}\}. \tag{61}$$

In addition, the change of variables (35), i.e. $U(\cdot, \theta) = B(\kappa(\theta), \cdot)$, is canonical for each $\theta \in \mathbb{T}^d$.

[15] Since the different references use different signs, we point out that here

$$\{F, G\} = \sum_j \left(\frac{\partial F}{\partial p_j} \frac{\partial G}{\partial q_j} - \frac{\partial F}{\partial q_j} \frac{\partial G}{\partial p_j} \right).$$

Proof: As noted before, we have that, for each $t \in \mathbb{R}$, $\alpha(t) = \gamma(t, t\omega) \in \widehat{\mathcal{G}}$. Application of Lemma 2.4 implies that, for all t and θ, $\gamma(t,\theta) \in \widehat{\mathcal{G}}$, and thus, according to (30)–(32), $\bar{\beta} \in \widehat{\mathfrak{g}}$, so that (60) holds.

Since volume-preserving maps form a group, divergence-free vector fields form a Lie algebra and commutators of divergence-free vector fields are again divergence free. The definition (33) proves that $\kappa(\theta)$ is volume preserving for each θ. This proves (ii).

A similar argument proves that in (iii) the averaged vector field and the change of variables are canonical. The formula (61) is a direct consequence of the standard homomorphism of the Lie algebra of Hamiltonian functions/Poisson brackets onto the Lie algebra of canonical vector fields/commutators [2].

□

Remark 3.3 Since the relations (54)–(55) are obviously nonlinear, the B-series obtained by taking the arithmetic mean of several B-series with coefficients in $\widehat{\mathcal{G}}$ does not have coefficients in $\widehat{\mathcal{G}}$, except in trivial cases. For this reason the coefficients (44) are such that $\chi, \chi^{-1} \notin \widehat{\mathcal{G}}$ and the corresponding averaging does not share the favorable geometric properties of quasi-stroboscopic averaging.

4 Computing the coefficients

Next we give simple recursions to obtain the coefficients of the B-series $B(\gamma(t,\theta), y)$ that are the key to compute the quasi-stroboscopically averaged system and the associated change of variables (see Table 2). Since, from the proof of Theorem 3.2 we know that $\gamma(t,\theta) \in \widehat{\mathcal{G}}$, it is enough to obtain the coefficients $\gamma_{\mathbf{k}_1\cdots\mathbf{k}_r}(t,\theta) := \gamma_{u_{\mathbf{k}_1\cdots\mathbf{k}_r}}(t,\theta)$, where the tall trees $u_{\mathbf{k}_1\cdots\mathbf{k}_r} \in \mathcal{T}$ were defined in the preceding subsection.[16] The notation $\mathbf{0}^r$ is shorthand for the word $\mathbf{0}\cdots\mathbf{0}$ (r letters), etc.

Proposition 4.1 *Given $r \geq 1$, $\mathbf{k} \in \mathbb{Z}^d - \{\mathbf{0}\}$, and $\mathbf{l}_1, \ldots, \mathbf{l}_s \in \mathbb{Z}^d$,*

$$\begin{aligned}
\gamma_{\mathbf{k}}(t,\theta) &= \frac{i}{\mathbf{k}\cdot\omega}(1 - e^{i\mathbf{k}\cdot\theta}),\\
\gamma_{\mathbf{0}^r}(t,\theta) &= t^r/r!,\\
\gamma_{\mathbf{0}^r\mathbf{k}}(t,\theta) &= \frac{i}{\mathbf{k}\cdot\omega}(\gamma_{\mathbf{0}^{r-1}\mathbf{k}}(t,\theta) - \gamma_{\mathbf{0}^r}(t,\theta)e^{i\mathbf{k}\cdot\theta}),\\
\gamma_{\mathbf{k}\mathbf{l}_1\cdots\mathbf{l}_s}(t,\theta) &= \frac{i}{\mathbf{k}\cdot\omega}(\gamma_{\mathbf{l}_1\cdots\mathbf{l}_s}(t,\theta) - \gamma_{(\mathbf{k}+\mathbf{l}_1)\mathbf{l}_2\cdots\mathbf{l}_s}(t,\theta)),\\
\gamma_{\mathbf{0}^r\mathbf{k}\mathbf{l}_1\cdots\mathbf{l}_s}(t,\theta) &= \frac{i}{\mathbf{k}\cdot\omega}(\gamma_{\mathbf{0}^{r-1}\mathbf{k}\mathbf{l}_1\cdots\mathbf{l}_s}(t,\theta) - \gamma_{\mathbf{0}^r(\mathbf{k}+\mathbf{l}_1)\mathbf{l}_2\cdots\mathbf{l}_s}(t,\theta)).
\end{aligned} \tag{62}$$

Proof: Particularizing (27) for $u = u_{\mathbf{k}_1\cdots\mathbf{k}_r}$, we see that $\gamma_{\mathbf{k}_1\cdots\mathbf{k}_r}(t,\theta) := \gamma_{u_{\mathbf{k}_1\cdots\mathbf{k}_r}}(t,\theta)$ satisfies

$$(\partial_t + \omega\cdot\nabla_\theta)\gamma_{\mathbf{k}_1\cdots\mathbf{k}_r}(t,\theta) = \gamma_{\mathbf{k}_1\cdots\mathbf{k}_{r-1}}(t,\theta)e^{i\mathbf{k}_r\cdot\theta}, \qquad \gamma_{\mathbf{k}_1\cdots\mathbf{k}_r}(0,\mathbf{0}) = 0. \tag{63}$$

Lemma 2.7 can be used to prove by induction on r that, for each $\mathbf{k}_1, \ldots, \mathbf{k}_r \in \mathbb{Z}^d$, (63) has a unique polynomial solution $\gamma_{\mathbf{k}_1\cdots\mathbf{k}_r}(t,\theta)$. Hence, the proposition follows if the coefficients $\gamma_{\mathbf{k}_1\cdots\mathbf{k}_r}(t,\theta)$ recursively determined by (62) satisfy (63), which we next prove by induction on r. One can trivially check that (63) is true for $r = 1$. Now, by applying the linear operator $(\partial_t + \omega\cdot\nabla_\theta)$ to both sides of the equality defining each $\gamma_{\mathbf{k}_1\cdots\mathbf{k}_r}(t,\theta)$ with $r > 1$ in (62) and applying the induction hypothesis, one arrives at (63). □

Proposition 4.1 provides the following recursive formulae to compute the coefficients $\bar{\beta}_{\mathbf{k}_1\cdots\mathbf{k}_r} = \partial_t\gamma_{\mathbf{k}_1\cdots\mathbf{k}_r}(0,\mathbf{0})$ required in the representation (60) of the averaged equations in quasi-stroboscopic averaging:

[16] The recursive formulae in Proposition 4.1 may be generalized to compute $\gamma_u(t,\theta)$ for arbitrary $u \in \mathcal{T}$. Alternatively, one can compute the coefficients $\gamma_u(t,\theta)$ for all u by applying the explicit formulae given in [19] to find the coefficients α_u ($u \in \mathcal{T}$) of an arbitrary $\alpha \in \widehat{\mathcal{G}}$ in terms of the corresponding $\alpha_{\mathbf{k}_1\cdots\mathbf{k}_r}$.

Theorem 4.2 *Given $r \geq 1$, $\mathbf{k} \in \mathbb{Z}^d \backslash \{0\}$, and $\mathbf{l}_1, \ldots, \mathbf{l}_s \in \mathbb{Z}^d$,*

$$\begin{aligned}
\bar{\beta}_{\mathbf{k}} &= 0, \\
\bar{\beta}_{\mathbf{0}} &= 1, \\
\bar{\beta}_{\mathbf{0}^{r+1}} &= 0, \\
\bar{\beta}_{\mathbf{0}^r\mathbf{k}} &= \frac{i}{\mathbf{k}\cdot\omega}(\bar{\beta}_{\mathbf{0}^{r-1}\mathbf{k}} - \bar{\beta}_{\mathbf{0}^r}), \\
\bar{\beta}_{\mathbf{k}\mathbf{l}_1\cdots\mathbf{l}_s} &= \frac{i}{\mathbf{k}\cdot\omega}(\bar{\beta}_{\mathbf{l}_1\cdots\mathbf{l}_s} - \bar{\beta}_{(\mathbf{k}+\mathbf{l}_1)\mathbf{l}_2\cdots\mathbf{l}_s}), \\
\bar{\beta}_{\mathbf{0}^r\mathbf{k}\mathbf{l}_1\cdots\mathbf{l}_s} &= \frac{i}{\mathbf{k}\cdot\omega}(\bar{\beta}_{\mathbf{0}^{r-1}\mathbf{k}\mathbf{l}_1\cdots\mathbf{l}_s} - \bar{\beta}_{\mathbf{0}^r(\mathbf{k}+\mathbf{l}_1)\mathbf{l}_2\cdots\mathbf{l}_s}).
\end{aligned}$$

To conclude this section, we shall give the first few terms in (60) more explicitly. By application of the Jacobi identity and the skew-symmetry of the commutators, (60) can be rewritten as

$$\begin{aligned}
\frac{d}{dt}Y &= \varepsilon\sum_{\mathbf{k}}\bar{\beta}_{\mathbf{k}}f_{\mathbf{k}} + \varepsilon^2\sum_{\mathbf{k}>\mathbf{l}}\bar{\beta}_{\mathbf{l}\mathbf{k}}[f_{\mathbf{l}}, f_{\mathbf{k}}] \\
&\quad + \varepsilon^3\left(\sum_{\mathbf{k}\neq\mathbf{l}}\bar{\beta}_{\mathbf{l}\mathbf{l}\mathbf{k}}[f_{\mathbf{l}}, [f_{\mathbf{l}}, f_{\mathbf{k}}]] + \sum_{\mathbf{l}>\mathbf{k}<\mathbf{m},\ \mathbf{m}\neq\mathbf{l}}\bar{\beta}_{\mathbf{m}\mathbf{l}\mathbf{k}}\,[f_{\mathbf{m}}, [f_{\mathbf{l}}, f_{\mathbf{k}}]]\right) + \mathcal{O}(\varepsilon^4),
\end{aligned}$$

where $<$ is some total ordering in the set of multi-indices $\mathbb{Z}^d$. We may assume that the ordering is such that $\mathbf{k} > \mathbf{0}$ for $\mathbf{k} \neq \mathbf{0}$. By taking into account the actual values of $\bar{\beta}_{\mathbf{k}_1\cdots\mathbf{k}_r}$ obtained from Theorem 4.2, we get the truncated averaged system

$$\frac{d}{dt}Y = \varepsilon f_{\mathbf{0}} + \varepsilon^2 F_2 + \varepsilon^3 F_3 + \mathcal{O}(\varepsilon^4), \tag{64}$$

where

$$\begin{aligned}
F_2 &= \sum_{\mathbf{k}>-\mathbf{k}}\frac{i}{\mathbf{k}\cdot\omega}([f_{\mathbf{k}} - f_{-\mathbf{k}}, f_{\mathbf{0}}] + [f_{-\mathbf{k}}, f_{\mathbf{k}}]), \\
F_3 &= \sum_{\mathbf{k}\neq\mathbf{0}}\frac{1}{(\mathbf{k}\cdot\omega)^2}\left([f_{\mathbf{0}}, [f_{\mathbf{0}}, f_{\mathbf{k}}]] + [f_{\mathbf{k}}, [f_{\mathbf{k}}, f_{-\mathbf{k}}]] - \frac{1}{2}[f_{\mathbf{k}}, [f_{\mathbf{k}}, f_{-2\mathbf{k}}]] + [f_{-\mathbf{k}}, [f_{\mathbf{k}}, f_{\mathbf{0}}]]\right) \\
&\quad + \sum_{\mathbf{0}\neq\mathbf{m}\neq-\mathbf{l}\neq\mathbf{0}}\frac{-1}{(\mathbf{l}\cdot\omega)((\mathbf{m}+\mathbf{l})\cdot\omega)}\,[f_{\mathbf{m}}, [f_{\mathbf{l}}, f_{\mathbf{0}}]] + \sum_{-\mathbf{l}>\mathbf{k}<\mathbf{l},\ \mathbf{k}\neq\mathbf{0}}\frac{1}{(\mathbf{k}\cdot\omega)(\mathbf{l}\cdot\omega)}\,[f_{-\mathbf{l}}, [f_{\mathbf{l}}, f_{\mathbf{k}}]] \\
&\quad + \sum_{\substack{\mathbf{m}>\mathbf{k}<-\mathbf{k}\\ \mathbf{m}+\mathbf{k}\neq\mathbf{0}}}\frac{-1}{(\mathbf{k}\cdot\omega)(\mathbf{m}\cdot\omega)}\,[f_{\mathbf{m}}, [f_{-\mathbf{k}}, f_{\mathbf{k}}]] + \sum_{\substack{\mathbf{0}\neq\mathbf{m}\neq\pm\mathbf{l}\neq\mathbf{0}\\ \mathbf{m}>-\mathbf{m}-\mathbf{l}<\mathbf{l}}}\frac{-1}{(\mathbf{m}\cdot\omega)((\mathbf{m}+\mathbf{l})\cdot\omega)}\,[f_{\mathbf{m}}, [f_{\mathbf{l}}, f_{-\mathbf{m}-\mathbf{l}}]].
\end{aligned}$$

When $d = 1$, formula (64) coincides with the messy formula (29) in [10]; a comparison of these two formulae clearly bears out the advantage of expressing the averaged system in terms of commutators as in (60).

5 Averaging for a class of near-integrable systems

In this section we study autonomous systems of the form

$$\frac{d}{dt}x = \sum_{j=1}^{d}\omega_j g_j(x) + \varepsilon\, h(x), \tag{65}$$

where g_j and h are smooth functions[17] in $\mathbb{R}^D$ and the flows $\Psi_t^{[j]}$ of the systems

$$\frac{d}{dt}x = g_j(x)$$

are assumed to be 2π-periodic and commute with each other, i.e. $\Psi_s^{[i]} \circ \Psi_t^{[j]} = \Psi_t^{[j]} \circ \Psi_s^{[i]}$. It is obvious that if, for $\theta = (\theta_1, \ldots, \theta_d) \in \mathbb{T}^d$, we set

$$\Psi_\theta := \Psi_{\theta_1}^{[1]} \circ \cdots \circ \Psi_{\theta_d}^{[d]},$$

then, for each $\theta, \theta' \in \mathbb{T}^d$, $\Psi_\theta \circ \Psi_{\theta'} = \Psi_{\theta+\theta'}$, so that the family $\{\Psi_\theta \ : \ \theta \in \mathbb{T}^d\}$ is a commutative d-parameter group of transformations in $\mathbb{R}^D$. As we shall discuss later, this setting caters for many near-integrable Hamiltonian systems.

The time-dependent change of variables $x = \Psi_{t\omega}(y)$ transforms (65) into

$$\frac{d}{dt}y = \varepsilon f(y, t\omega), \qquad f(y,\theta) := \left(\frac{\partial}{\partial y}\Psi_\theta(y)\right)^{-1} h(\Psi_\theta(y)), \tag{66}$$

a system of the format (1) considered in the preceding sections.

Remark 5.1 When the non-autonomous system (1) is rewritten as

$$\frac{d}{dt}y = \varepsilon f(y,\theta), \qquad \frac{d}{dt}\theta = \omega,$$

it becomes a member of the class (65) with $\Psi_t^{[j]}(y,\theta) = (y, \theta + t\mathbf{e}_j)$ ($\mathbf{e}_j$ denotes the j-th unit vector in $\mathbb{R}^d$). Since, as we have just seen, the autonomous (65) may be recast in non-autonomous form, we conclude that (1) and (65) may be regarded as equivalent.

The application of Theorem 2.10 to (66) shows that

$$y(t) = B(\gamma(t, t\omega), y(0)),$$

where, of course, γ is the map we found before and the elementary differentials implied in the B-series are those associated with the function f defined in (66). As a consequence, the solution flow of (65) may be represented as

$$x(t) = \Psi_{t\omega}(B(\gamma(t, t\omega), x(0)));$$

this is the starting point to construct an averaged system to describe the dynamics of $x(t)$.

5.1 The averaged system

If, for each $t \in \mathbb{R}$, $\theta \in \mathbb{T}^d$, we define

$$\Phi_{t,\theta}(x) := \Psi_\theta(B(\gamma(t,\theta), x)), \tag{67}$$

then the flow of (65) is given by

$$x(t) = \Phi_{t,t\omega}(x(0)). \tag{68}$$

In order to study the properties of the mapping $\Phi_{t,\theta}$, we first require the following:

Lemma 5.2 *For any $\delta \in \widehat{\mathcal{G}}$, $\theta \in \mathbb{T}^d$ and $x \in \mathbb{R}^D$,*

$$B(\delta, \Psi_\theta(x)) = \Psi_\theta(B(\delta^\theta, x)) \tag{69}$$

($\delta^\theta \in \widehat{\mathcal{G}}$ was defined in (51)).

[17] More precisely it is required that g_j and h are such that the transformed vector field f in (66) satisfies the requirements specified at the beginning of Section 2.

Proof: Consider, for fixed $\theta \in \mathbb{T}^d$, the change of variables $y = \Psi_\theta(x)$. From Proposition 3.1 we know that

$$B(\delta, \Psi_\theta(x)) = \Psi_\theta(\tilde{B}(\delta, x)),$$

where the notation $\tilde{B}$ means that the implied elementary differentials $\tilde{\mathcal{F}}_u(x)$ are built from the pulled-back vector fields

$$\tilde{f}_{\mathbf{k}}(x) := \left(\frac{\partial}{\partial x}\Psi_\theta(x)\right)^{-1} f_{\mathbf{k}}(\Psi_\theta(x))$$

instead of from $f_{\mathbf{k}}(y)$. Thus, we have to prove that $\tilde{B}(\delta, x) = B(\delta^\theta, x)$ and in order to do so it is sufficient to establish that, for each tree u,

$$\tilde{\mathcal{F}}_u(x) = e^{i\mathcal{I}(u)\cdot\theta}\mathcal{F}_u(x),$$

a fact that is a consequence of the equality $\tilde{f}_{\mathbf{k}}(x) = e^{i\mathbf{k}\cdot\theta} f_{\mathbf{k}}(x)$ which we shall prove next.

The definition of $f_{\mathbf{k}}$ as a Fourier coefficient of f in (66) yields:

$$f_{\mathbf{k}}(y) = \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} e^{-i\mathbf{k}\cdot\theta'} f(y,\theta')\,d\theta' = \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} e^{-i\mathbf{k}\cdot\theta'}\left(\frac{\partial}{\partial y}\Psi_{\theta'}(y)\right)^{-1} h(\Psi_{\theta'}(y))\,d\theta'.$$

Then, using the definition of the pulled-back $\tilde{f}_{\mathbf{k}}(x)$ above and the group property of Ψ_θ, we may proceed as follows:

$$\begin{aligned}
\tilde{f}_{\mathbf{k}}(x) &= \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} e^{-i\mathbf{k}\cdot\theta'}\left(\frac{\partial}{\partial x}\Psi_\theta(x)\right)^{-1}\left(\frac{\partial}{\partial y}\Psi_{\theta'}(y)\right)^{-1} h(\Psi_{\theta'}(y))\,d\theta' \\
&= \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} e^{-i\mathbf{k}\cdot\theta'}\left(\frac{\partial}{\partial x}\Psi_{\theta+\theta'}(x)\right)^{-1} h(\Psi_{\theta+\theta'}(x))d\theta' \\
&= \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} e^{-i\mathbf{k}\cdot\theta'} f(x,\theta+\theta')d\theta' \\
&= \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} e^{-i\mathbf{k}\cdot(\theta'-\theta)} f(x,\theta')d\theta' \\
&= e^{i\mathbf{k}\cdot\theta} f_{\mathbf{k}}(x).
\end{aligned}$$

□

We are now in a position to prove that the collection $\{\Phi_{t,\theta}\}$ is a $(d+1)$-parameter commutative group:

Proposition 5.3 *For each $t, t' \in \mathbb{R}$ and $\theta, \theta' \in \mathbb{T}^d$:*

$$\Phi_{t',\theta'} \circ \Phi_{t,\theta} = \Phi_{t+t',\theta+\theta'}.$$

Proof: By virtue of Lemma 5.2, we have that

$$\begin{aligned}
\Phi_{t',\theta'} \circ \Phi_{t,\theta} &= \Psi_{\theta'} \circ B(\gamma(t',\theta'),\cdot) \circ \Psi_\theta \circ B(\gamma(t,\theta),\cdot) \\
&= \Psi_{\theta'} \circ \Psi_\theta \circ B(\gamma^\theta(t',\theta'),\cdot) \circ B(\gamma(t,\theta),\cdot) \\
&= \Psi_{\theta+\theta'} \circ B(\gamma(t,\theta) * \gamma^\theta(t',\theta'),\cdot),
\end{aligned}$$

and, by (53) and (28)–(29), we may write

$$\begin{aligned}
\gamma(t,\theta) * \gamma^\theta(t',\theta') &= \gamma(t,\mathbf{0}) * \gamma(0,\theta) * \gamma^\theta(t',\theta') \\
&= \gamma(t,\mathbf{0}) * \gamma(t',\theta+\theta') \\
&= \gamma(t+t',\theta+\theta').
\end{aligned}$$

□

It follows from the proposition that the flow (68) of the original system (65) admits the factorizations

$$\Phi_{t,t\omega} = \Phi_{0,t\omega} \circ \Phi_{t,\mathbf{0}} = \Phi_{t,\mathbf{0}} \circ \Phi_{0,t\omega}; \tag{70}$$

here (see (33))

$$\Phi_{0,\theta}(\cdot) = \Psi_\theta\big(B(\gamma(0,\theta),\cdot)\big) = \Psi_\theta\big(B(\kappa(\theta),\cdot)\big) \tag{71}$$

is a map parameterized by $\theta \in \mathbb{T}^d$, and

$$\Phi_{t,\mathbf{0}}(\cdot) = \Psi_{\mathbf{0}}\big(B(\gamma(t,\mathbf{0}),\cdot)\big) = B(\bar\alpha,\cdot)$$

is the t-flow (see (31)–(32)) of an autonomous (averaged) system

$$\frac{d}{dt}X = \varepsilon\,\tilde h(X), \qquad \varepsilon\,\tilde h(X) := \frac{d}{dt}\Phi_{t,\mathbf{0}}(X)\bigg|_{t=0} = B(\bar\beta, X). \tag{72}$$

The factorizations (70) lead to the following result.

Theorem 5.4 *(i) Let x be a solution of (65). Then, $x(t) = \Phi_{0,t\omega}(X(t))$, where X is the solution of the averaged system (72) with initial value $X(0) = x(0)$, and $\Phi_{0,\theta}$ is the map in (71).*

Alternatively, for each $t \in \mathbb{R}$, $x(t) = X(t)$, where X is the solution of the averaged system (72) with initial value $X(0) = \Phi_{0,t\omega}(x(0))$.

(ii) Assume that $\Lambda : \mathbb{R}^D \to \mathbb{R}$ is a first integral of the given system (65), i.e. $\Lambda(x(t)) = \Lambda(x(0))$ for each $t \in \mathbb{R}$ and each solution x. Then Λ is a first integral of the averaged system (72), $\Lambda(X(t)) \equiv \Lambda(X(0))$. Furthermore, for each $\theta \in \mathbb{T}^d$, the transformation $\Phi_{0,\theta}$ also preserves the value of Λ: $\Lambda \circ \Phi_{0,\theta} = \Lambda$.

Proof: (i) is a direct consequence of (70). For (ii) we note that $\Lambda\circ\Phi_{t,t\omega} = \Lambda$ implies $\Lambda\circ\Phi_{t,\mathbf{0}} = \Lambda$ and $\Lambda\circ\Phi_{0,\theta} = \Lambda$; this is proved by expanding $\Lambda(\Phi_{t,\theta}(\cdot)) - \Lambda(\cdot)$ in powers of ε and applying Lemma 2.4 at each order of ε. □

It is perhaps useful to recall that $\varepsilon\,\tilde h(X) = B(\bar\beta, X)$ may be written as in the right-hand side of (60) (with Y replaced by X), where the coefficients $\bar\beta_{\mathbf{k}_1\cdots\mathbf{k}_r}$ may be recursively obtained from Theorem 4.2.

5.2 A decomposition of the vector field

The group property of the maps $\Phi_{t,\theta}$ implies that the right-hand side of (65) may be re-written as the sum of $d+1$ vector fields that commute with each other; the details are as follows. For each j, the maps $\Phi_t^{[j]} := \Phi_{0,t\mathbf{e}_j}$ form a one-parameter semigroup and therefore are the t-flow of the autonomous system

$$\frac{d}{dt}x = \tilde g_j(x), \qquad \tilde g_j(x) := \frac{d}{dt}\Phi_{0,t\mathbf{e}_j}(x)\bigg|_{t=0}. \tag{73}$$

(Note that $\Phi_t^{[j]}$ depends 2π-periodically on t.) Differentiation in (67) reveals that the fields $\tilde g_j$ are related to the fields g_j by

$$\tilde g_j(x) = g_j(x) + B(\beta^{[j]}, x), \qquad \beta^{[j]} := \frac{d}{dt}\gamma(0,t\mathbf{e}_j)\bigg|_{t=0} = \frac{d}{dt}\kappa(t\mathbf{e}_j)\bigg|_{t=0}. \tag{74}$$

The factorizations in (70) may now be taken further:

$$\Phi_{0,t\omega} = \Phi^{[1]}_{t\omega_1} \circ \cdots \circ \Phi^{[d]}_{t\omega_d}, \qquad \Phi_{t,t\omega} = \Phi_{t,\mathbf{0}} \circ \Phi^{[1]}_{t\omega_1} \circ \cdots \circ \Phi^{[d]}_{t\omega_d}.$$

(The flows in the right-hand sides of these identities commute with each other.) Accordingly Theorem 5.4 may be supplemented by the following result:

Theorem 5.5 *(i) The map $\Phi_{0,t\omega}$ in Theorem 5.4 is the time t flow of the system*

$$\frac{d}{dt}x = \sum_{j=1}^{d} \omega_j \tilde{g}_j(x),$$

where the fields $\tilde{g}_j(x)$ are given by (74) and generate 2π-periodic flows.

(ii) The given system (65) may be rewritten as

$$\frac{d}{dt}x = \sum_{j=1}^{d} \omega_j \tilde{g}_j(x) + \varepsilon\, \tilde{h}(x),$$

where each of the fields $\omega_j \tilde{g}_j(x)$, $j = 1, \ldots, d$, $\tilde{h}(x)$ commute with each other and $\varepsilon\tilde{h}$ is the averaged vector field (72).

(iii) Assume that Λ is a first integral of the given system (65). Then Λ is also a first integral of each of the systems (73).

Remark 5.6 Theorem 5.5 is related to the application of classical normal form theory in the semisimple case with imaginary eigenvalues [22] as follows: If each $g_j(x)$ is linear , then (65) is of the form $dx/dt = Ax + \varepsilon h(x)$, where A is a diagonalizable matrix with imaginary eigenvalues that are linear combinations with integer coefficients of the frequencies $\omega_1, \ldots, \omega_d$. In normal form theory, a near-identity formal change of variables is recursively constructed such that in the new variables z, (65) reads $dz/dt = Az + \varepsilon \hat{h}(z)$, where $\hat{h}(z)$ commutes with Az. In the original variables, this gives a decomposition of (65) of the form provided by Theorem 5.5.

In order to compute explicitly the $\tilde{g}_j(x)$, we note that, since $\gamma(t,\theta) \in \widehat{\mathcal{G}}$ for each t and θ, we have $\beta^{[j]} \in \widehat{\mathfrak{g}}$ and thus

$$\tilde{g}_j(x) = g_j(x) + \sum_{r=1}^{\infty} \sum_{\mathbf{k}_1,\ldots,\mathbf{k}_r \in Z^d} \frac{\varepsilon^r}{r}\, \beta^{[j]}_{\mathbf{k}_1\cdots\mathbf{k}_r}\, [[\cdots[[f_{\mathbf{k}_1}, f_{\mathbf{k}_2}], f_{\mathbf{k}_3}]\cdots], f_{\mathbf{k}_r}](x), \tag{75}$$

where

$$\beta^{[j]}_{\mathbf{k}_1\cdots\mathbf{k}_r} := \left.\frac{d}{dt}\gamma_{\mathbf{k}_1\cdots\mathbf{k}_r}(0, t\mathbf{e}_j)\right|_{t=0}.$$

In turn, the coefficients $\gamma_{\mathbf{k}_1\cdots\mathbf{k}_r}(t,\theta)$ are given recursively in Proposition 4.1. The next result, a direct consequence of that proposition, lists the recursive formulae to compute the $\beta^{[j]}$.

Theorem 5.7 *Given $r \geq 1$, $\mathbf{k} \in \mathbb{Z}^d \backslash \{0\}$, and $\mathbf{l}_1, \ldots, \mathbf{l}_s \in \mathbb{Z}^d$,*

$$\begin{aligned}
\beta^{[j]}_{\mathbf{k}} &= \frac{\mathbf{k}\cdot\mathbf{e}_j}{\mathbf{k}\cdot\omega}, \\
\beta^{[j]}_{\mathbf{0^r}} &= 0, \\
\beta^{[j]}_{\mathbf{0}^r\mathbf{k}} &= \frac{i}{\mathbf{k}\cdot\omega}(\beta^{[j]}_{\mathbf{0}^{r-1}\mathbf{k}} - \beta^{[j]}_{\mathbf{0}^r}), \\
\beta^{[j]}_{\mathbf{k}\mathbf{l}_1\cdots\mathbf{l}_s} &= \frac{i}{\mathbf{k}\cdot\omega}(\beta^{[j]}_{\mathbf{l}_1\cdots\mathbf{l}_s} - \beta^{[j]}_{(\mathbf{k}+\mathbf{l}_1)\mathbf{l}_2\cdots\mathbf{l}_s}), \\
\beta^{[j]}_{\mathbf{0}^r\mathbf{k}\mathbf{l}_1\cdots\mathbf{l}_s} &= \frac{i}{\mathbf{k}\cdot\omega}(\beta^{[j]}_{\mathbf{0}^{r-1}\mathbf{k}\mathbf{l}_1\cdots\mathbf{l}_s} - \beta^{[j]}_{\mathbf{0}^r(\mathbf{k}+\mathbf{l}_1)\mathbf{l}_2\cdots\mathbf{l}_s}).
\end{aligned}$$

5.3 The Hamiltonian case

Let us assume hereafter that D is even and that the fields $g_j(x)$, $j = 1, \ldots, d$, and $h(x)$ in (65) are all canonical and arise from the Hamiltonian functions *in involution* $I_j(x)$, $j = 1, \ldots, d$, and $K(x)$ respectively. Then (65) is itself a canonical system with Hamiltonian function

$$\mathcal{H}(x) := \sum_{j=1}^{d} \omega_j I_j(x) + \varepsilon K(x). \tag{76}$$

In the unperturbed, $\varepsilon = 0$, case, the hypothesis that the flows $\Psi_t^{[j]}$ commute with each other imply that the I_j are d conserved quantities of (65) in involution [2]; then (65) is integrable with flow given by the composition $\Psi_{t\omega_1}^{[1]} \circ \cdots \circ \Psi_{t\omega_d}^{[d]}$. Thus $\mathcal{H}$ is near-integrable, and in fact many near-integrable Hamiltonians may be brought to the form (76); some examples follow.

- Assume that $x = (a, \theta)$, with $a \in \mathbb{R}^{D/2}$ and $\theta \in \mathbb{T}^{D/2}$, and that

  $$\mathcal{H}(x) = \sum_{j=1}^{D/2} \lambda_j a_j + \varepsilon K(a, \theta).$$

  Here a and θ are action and angle variables in the unperturbed problem and the λ_j provide the $D/2$, possibly resonant, corresponding angular frequencies. We may find a non-resonant $\omega \in \mathbb{R}^d$, with $d \leq D/2$ such that each λ_j is a linear combination with integer coefficients of the ω_k and then $\mathcal{H}$ takes the form (76).

- Assume that

  $$\mathcal{H} = H_0(x) + \varepsilon K(x), \tag{77}$$

  where H_0 is quadratic and possesses $D/2$ oscillatory normal modes with frequencies $\lambda_1, \ldots, \lambda_{D/2}$. Again the λ_j may be rewritten in terms of non-resonant ω_k and then $\mathcal{H}$ takes the form (76) with quadratic I_j's.

In the application of the material of the preceding subsections to the present Hamiltonian case, we note that the transformation Ψ_θ is canonical for each $\theta \in \mathbb{T}^d$ and this implies that the transformed system (66) is also canonical. Theorem 3.2 (iii) then reveals that the averaged system (72) is canonical with a Hamiltonian function $\varepsilon \tilde{K}$ given by the right-hand side of (61), where the $H_{\mathbf{k}}$ are the Fourier coefficients of the Hamiltonian $K(\Psi_\theta(y))$ of the transformed field in (66). In addition we have the following result:

Theorem 5.8 *(i) Each system (73) is canonical, with Hamiltonian function*

$$\tilde{I}_j(x) := I_j(x) + \sum_{r=1}^{\infty} \sum_{\mathbf{k}_1, \ldots, \mathbf{k}_r \in Z^d} \frac{\varepsilon^r}{r} \beta^{[j]}_{\mathbf{k}_1 \cdots \mathbf{k}_r} \{\{\cdots\{\{H_{\mathbf{k}_1}, H_{\mathbf{k}_2}\}, H_{\mathbf{k}_3}\}\cdots\}, H_{\mathbf{k}_r}\}(x).$$

(ii) The functions $\tilde{I}_j(x)$, $j = 1, \ldots, d$, are first integrals in involution of both the given system (65) and of the averaged system (72).

Proof: Part (i) is a consequence of the fact that $\beta^{[j]}$ belongs to $\widehat{\mathfrak{g}}$. For part (ii) we note that $\mathcal{H}$ is of course a first integral of (65); then Theorem 5.5 (iii) implies that $\mathcal{H}$ is a first integral of each of the systems (73) or, in other words, that the Poisson brackets $\{\mathcal{H}, \tilde{I}_j\}$ vanish. This in turn shows that each $\tilde{I}_j$ is a first integral of the original system with Hamiltonian $\mathcal{H}$. Theorem 5.4 (ii) yields finally that each $\tilde{I}_j$ is a first integral of (72). □

The first terms in the expansion of $\tilde{I}_j(x)$ are explicitly given by

$$\begin{aligned}\tilde{I}_j(x) &= I_j(x) + \varepsilon \sum_{\mathbf{k}>\mathbf{0}} \frac{\mathbf{k}\cdot\mathbf{e}_j}{\mathbf{k}\cdot\omega}\,(H_{\mathbf{k}}(x) + H_{-\mathbf{k}}(x)) \\ &\quad + \varepsilon^2 \sum_{\mathbf{k}>\mathbf{0}} i\,\frac{\mathbf{k}\cdot\mathbf{e}_j}{(\mathbf{k}\cdot\omega)^2}\,(\{H_{\mathbf{0}}, H_{\mathbf{k}} - H_{-\mathbf{k}}\}(x) + \{H_{\mathbf{k}}, H_{-\mathbf{k}}\}(x)) \\ &\quad + \varepsilon^2 \sum_{0<\mathbf{k}<\mathbf{l}} (A_{\mathbf{k},\mathbf{l},j}(x) + A_{-\mathbf{k},\mathbf{l},j}(x) + A_{\mathbf{k},-\mathbf{l},j}(x) + A_{-\mathbf{k},-\mathbf{l},j}(x)) + \mathcal{O}(\varepsilon^3),\end{aligned} \tag{78}$$

where

$$A_{\mathbf{k},\mathbf{l},j}(x) = \frac{i}{(\mathbf{l}+\mathbf{k})\cdot\omega}\Big(\frac{\mathbf{k}\cdot\mathbf{e}_j}{\mathbf{k}\cdot\omega} - \frac{\mathbf{l}\cdot\mathbf{e}_j}{\mathbf{l}\cdot\omega}\Big)\{H_{\mathbf{l}}, H_{\mathbf{k}}\}(x).$$

5.4 An application

We consider the Hamiltonian system with five degrees of freedom ($q \in \mathbb{R}^5$, $p \in \mathbb{R}^5$)

$$\begin{aligned}H(p,q) &= \lambda_1\left(\frac{1}{2}p_1^2 + \frac{1}{2}q_1^2\right) + \sum_{j=2}^{5}\left(\frac{1}{2}p_j^2 + \frac{\lambda_j^2}{2}q_j^2\right) + \varepsilon\, U(q), \\ U(q) &= \frac{\delta^2}{8}q_1^2q_2^2 + \delta^4\left(\frac{\sqrt{70}}{20} + q_2 + q_3 + \frac{5}{2}q_4 + q_5\right)^4,\end{aligned} \tag{79}$$

where

$$\lambda_1 = \varepsilon, \quad \lambda_2 = \lambda_3 = 1, \quad \lambda_4 = 2, \quad \lambda_5 = \sqrt{2}.$$

The specific case with parameter values $\varepsilon = 1/70$, $\delta = 1/\sqrt{70}$, and initial values

$$p(0) = (-0.2, 0.6\delta, 0.7\delta, 0.8\delta, -0.9\delta), \qquad q(0) = (1, 0.3\delta, 0.8\delta, 0.7\delta, -1.1\delta)$$

is considered in [13], XIII.9.1 (but variables there are scaled differently, in particular t in that reference corresponds to εt here).

In (79) the quadratic part corresponds to five uncoupled harmonic oscillators (cf. (77)) with frequencies $\lambda_1, \ldots, \lambda_5$, and Hamiltonian functions

$$J_1 = \lambda_1\left(\frac{1}{2}p_1^2 + \frac{1}{2}q_1^2\right), \qquad J_j = \frac{1}{2}p_j^2 + \frac{\lambda_j^2}{2}q_j^2, \quad j = 2, \ldots, 5.$$

In Figure 1 we have plotted, for the parameter values and initial conditions mentioned above, the evolution of the five quantities J_j and the sum $J_2 + J_3 + J_4$ for $0 \le t \le 500/\varepsilon^2$. We observe that J_1, $J_2 + J_3 + J_4$ and J_5 are approximately conserved over the whole time interval. (The conservation of J_1 is not studied in [13].)

To finish this section, we apply to (79) the results we have obtained in the article. The Hamiltonian (79), where for the time being δ is regarded as a constant whose value is independent of the value of the parameter ε, is of the form (76), with $x = (p,q)$, $\omega = (1, \sqrt{2}) \in \mathbb{R}^2$, and

$$I_1 = J_2 + J_3 + J_4, \qquad I_2 = \frac{1}{\sqrt{2}}J_5, \qquad K = \frac{1}{2}p_1^2 + \frac{1}{2}q_1^2 + U(q),$$

so that Theorems 5.5 and 5.8 lead to the decomposition

$$H(p,q) = \tilde{I}_1(p,q) + \sqrt{2}\tilde{I}_2(p,q) + \varepsilon\tilde{K}(p,q),$$

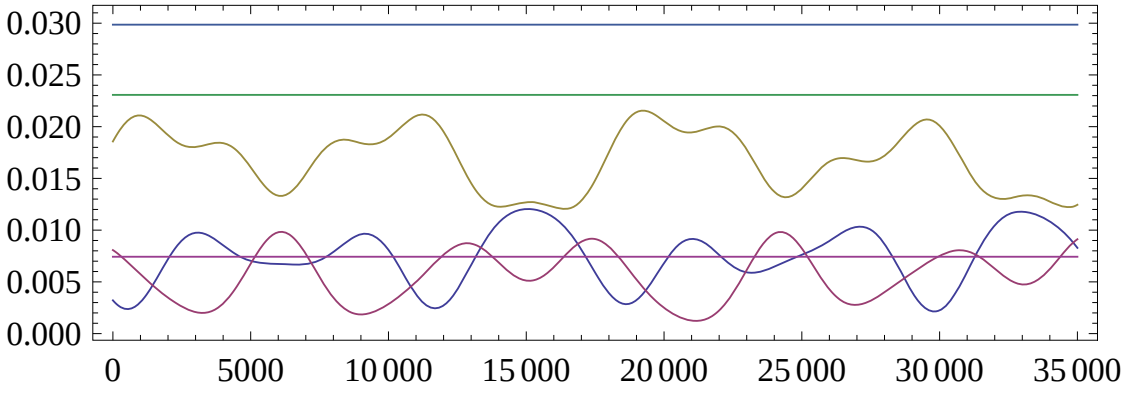


Figure 1: Energy exchange between J_2, J_3, J_4 (wiggly lines) and almost conservation of J_1 (bottom), J_5, $J_2 + J_3 + J_4$ (top); the horizontal axis corresponds to the variable εt.

with formal first integrals $\tilde{I}_1(p,q)$ and $\tilde{I}_2(p,q)$ in involution. These integrals correspond to the near-conservation of $J_2 + J_3 + J_4$ and J_5 along motions of (79). Moreover, the averaged Hamiltonian $\tilde{K}(p,q)$ happens to be of the form

$$\tilde{K}(p,q) = \frac{1}{2}p_1^2 + \frac{1}{2}q_1^2 + \delta^2\, R(p,q;\delta^2),$$

and therefore the averaging process can be applied to $\tilde{K}(p,q)$ with δ^2 considered as the perturbation parameter. This gives rise to a further decomposition

$$\tilde{K}(p,q) = \tilde{I}_3(p,q) + \delta^2\, \tilde{R}(p,q;\delta^2),$$

where $\tilde{I}_3(p,q) = \frac{1}{2}p_1^2 + \frac{1}{2}q_1^2 + \mathcal{O}(\delta^2)$ and $\tilde{I}_3(p,q)$ and $\tilde{K}(p,q)$ are in involution. Since H is a first integral of $\tilde{K}$, then $\tilde{I}_3$ is also a first integral of H. We then end up with a decomposition

$$H(p,q) = \tilde{I}_1(p,q) + \sqrt{2}\tilde{I}_2(p,q) + \varepsilon\tilde{I}_3(p,q) + \varepsilon\delta^2\tilde{R}(p,q;\delta^2),$$

where $\tilde{I}_j(p,q)$, $j = 1,2,3$, are in involution with $H(p,q)$. This matches the approximate conservation of $J_2 + J_3 + J_4$, J_5 and J_1 in Figure 1.

We have computed the formal first integrals $\tilde{I}_j(p,q)$, $j = 1,2,3$ (except for remainders of size $\mathcal{O}(\varepsilon^3)$, $\mathcal{O}(\varepsilon^3)$, and $\mathcal{O}(\delta^6)$ respectively) by applying formula (78); this results in polynomials in the variables p, q that consist of the sum of 1653, 1574 and 17050 monomials respectively (a computer algebra programme was of course used to compute the Poisson brackets that feature in (78)). In order to check the size of the variation of the truncated versions of $\tilde{I}_j(p,q)$ along the solution of the original Hamiltonian system, we have computed, in the interval $0 \le t \le 500/\varepsilon^2$ with a high-order splitting algorithm with small step size, the solution corresponding to the above initial data. In Figure 5.4, $(\tilde{I}_j(p(t),q(t)) - \tilde{I}_j(p(0),q(0))/\varepsilon^{d_j}$, with[18] $d_1 = d_3 = 5$, $d_2 = 7$, are plotted as functions of $\varepsilon^2 t$ for $\varepsilon = \delta^2 = 1/70$ (in blue) and $\varepsilon = \delta^2 = 1/140$ (in red) in respectively.

Appendix: multiplication of coefficients

In this appendix we give details of the product $*$ in $\mathcal{G}$. It is hoped that this will increase the readability of the paper; for a more complete treatment the reader is referred to [13], Chapter III.

[18]The exponent d_j in the scaling factor for each conserved quantity is determined by the scaling of the initial condition in tandem by the degree, as polynomials in the solution components, of the Poisson brackets discarded when truncating.

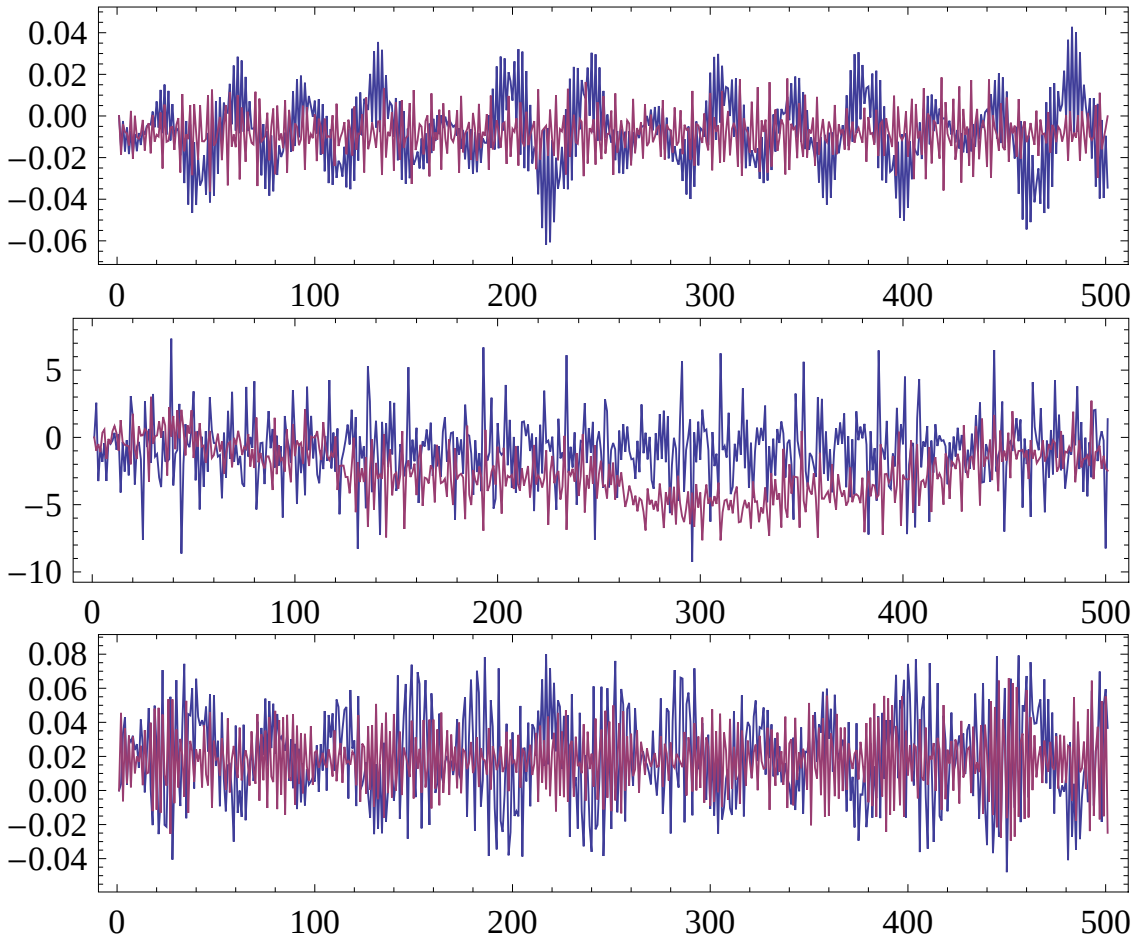


Figure 2: $(\tilde{I}_1(p(t),q(t)) - \tilde{I}_1(p(0),q(0)))/\varepsilon^5$, $(\tilde{I}_2(p(t),q(t)) - \tilde{I}_2(p(0),q(0))/\varepsilon^7$, $(\tilde{I}_3(p(t),q(t)) - \tilde{I}_3(p(0),q(0))/\varepsilon^5$ as functions of the scaled time $\varepsilon^2 t$. The blue solutions correspond to $\varepsilon = \delta^2 = 1/70$ and the red solutions to $\varepsilon = \delta^2 = 1/140$

Given $\delta \in \mathbb{C}^{\mathcal{T}\cup\{\emptyset\}}$ with $\delta_\emptyset = 1$ (that is $\delta \in \mathcal{G}$, so that $B(\delta, y)$ is a near-identity B-series), and an arbitrary B-series $B(\eta, y)$, the composition $B(\eta, B(\delta, y))$ is a B-series of the form $B(\zeta, y)$, where $\zeta_\emptyset = \eta_\emptyset\delta_\emptyset$ and for $u \in \mathcal{T}$ the coefficient $\zeta_u = (\delta * \eta)_u$, $u \in \mathcal{T}$ equals $\eta_\emptyset\delta_u + \eta_u\delta_\emptyset = \eta_\emptyset\delta_u + \eta_u$ plus a sum of products of the form $\eta_{u_0}\delta_{u_1}\cdots\delta_{u_m}$ with $|u_0| + |u_1| + \cdots + |u_m| = |u|$, $u_0, \ldots, u_m \in \mathcal{T}$. For instance, for the tree in the last row of Table 1 it holds that (we use brackets rather than subscripts for typographical convenience):

$$\zeta\Big(\begin{smallmatrix} & \mathbf{l} & & \mathbf{m} & \\ & & \mathbf{k} & & \end{smallmatrix}\Big) = \eta(\emptyset)\,\delta\Big(\begin{smallmatrix} & \mathbf{l} & & \mathbf{m} & \\ & & \mathbf{k} & & \end{smallmatrix}\Big) + \eta\big(\mathbf{k}\big)\,\delta\big(\mathbf{l}\big)\,\delta\big(\mathbf{m}\big) \tag{80}$$
$$+ \eta\Big(\begin{smallmatrix} \mathbf{l} \\ \mathbf{k} \end{smallmatrix}\Big)\,\delta\big(\mathbf{m}\big) + \eta\Big(\begin{smallmatrix} \mathbf{m} \\ \mathbf{k} \end{smallmatrix}\Big)\,\delta\big(\mathbf{l}\big) + \eta\Big(\begin{smallmatrix} & \mathbf{l} & & \mathbf{m} & \\ & & \mathbf{k} & & \end{smallmatrix}\Big)\,\delta(\emptyset).$$

Multiplication of the formula (80) by $\exp(i(\mathbf{k}+\mathbf{l}+\mathbf{m})\cdot\theta)$ establishes the validity of (52) at $u = \begin{smallmatrix} & \mathbf{l} & & \mathbf{m} & \\ & & \mathbf{k} & & \end{smallmatrix}$.

In the general case, the coefficient $(\delta * \eta)_u$, $u \in \mathcal{T}$ can be written as

$$(\delta * \eta)_u = \eta(L_\delta(u)), \tag{81}$$

where, for each fixed $\delta \in \mathcal{G}$, $L_\delta : \mathcal{V} \to \mathcal{V}$ is a linear map (defined below) in the vector space $\mathcal{V} = \mathrm{span}(\mathcal{T} \cup \{\emptyset\})$ of linear combinations (with complex coefficients) of rooted trees, and η (originally defined as a map $\mathcal{T} \cup \{\emptyset\} \to \mathbb{C}$) has been extended by linearity to $\mathcal{V}$ (i.e. $\eta(\sum_j \lambda_j u_j) = \sum_j \lambda_j \eta_{u_j}$ for each linear combination $\sum_j \lambda_j u_j$ of trees).

In order to define the linear map L_δ, we first extend the notation $[u_1 \cdots u_m]_{\mathbf{k}}$ (used to recursively represent trees in $\mathcal{T}$), to define for each $m \geq 1$ and each multi-index $\mathbf{k} \in \mathbb{Z}^d$, an m-linear, commutative m-ary operation on $\mathcal{V}$ by setting

$$[\emptyset]_{\mathbf{k}} = \textcircled{\scriptsize \mathbf{k}}, \qquad [u_1 \cdots u_{m-1} \emptyset]_{\mathbf{k}} = [u_1 \cdots u_{m-1}]_{\mathbf{k}}.$$

Now, $L_\delta(u)$ is defined recursively by

$$L_\delta(\emptyset) = \emptyset, \quad L_\delta(u) = \delta_u \, \emptyset + [L_\delta(u_1) \cdots L_\delta(u_m)]_{\mathbf{k}} \quad \text{if} \quad u = [u_1 \cdots u_m]_{\mathbf{k}}. \tag{82}$$

One can check that for any $u \in \mathcal{T}$, $L_\delta(u) = \delta(u) \, \emptyset + u + v$ where v a sum of rooted trees with fewer vertices than u.

Observe that pictorially, in the right-hand side of (80), the tree associated with the right factor η ranges in the set of trees obtained by pruning some edges in the tree in the left-hand side ('pruning' includes complete uprooting, as in the first term, or no pruning at all, as in the last). In each term, the trees associated with the left factor δ are precisely the parts that have been chopped off. It can be shown that (81)–(82) imply that the same pruning procedure is valid to write the formula for $(\delta * \eta)_u$ for arbitrary $u \in \mathcal{T}$.

Acknowledgement. P. Chartier has received financial support from INRIA through the associated team MIMOL. A. Murua and J.M. Sanz-Serna have been supported by projects MTM2010-18246-C03-03 and MTM2010-18246-C03-01 respectively from Ministerio de Ciencia e Innovación.

References

[1] V. I. Arnold, Geometrical Methods in the Theory of Ordinary Differential Equations, 2nd ed., Springer, New York, 1988.

[2] V. I. Arnold, Mathematical Methods of Classical Mechanics, 2nd ed., Springer, New York, 1989.

[3] J.C. Butcher, An algebraic theory of integration methods, *Math. Comput.* **26**, 79–106 (1972).

[4] Ch. Brouder, Trees, renormalization and differential equations, *BIT* **44**, 425–438 (2004).

[5] M.P. Calvo, Ph. Chartier, A. Murua, J.M. Sanz-Serna, A stroboscopic method for highly oscillatory problems. In: B. Engquist, O. Runborg, R. Tsai, (eds.) Numerical Analysis of Multiscale Computations, Lect. Notes Comput. Sci. Eng. 82, pp. 73–87. Springer, Berlin, 2011.

[6] M.P. Calvo, Ph. Chartier, A. Murua, J.M. Sanz-Serna, Numerical stroboscopic averaging for ODEs and DAEs, Applied Numerical Mathematics, *Appl. Numer. Math.* **61**, 1077–1095 (2011).

[7] M.P. Calvo, A. Murua, J.M. Sanz-Serna, Modified equations for ODEs, *Contemp. Math.* **172**, 63–74 (1994).

[8] P. Chartier, E. Faou, A. Murua, An algebraic approach to invariant preserving integrators: the case of quadratic and Hamiltonian invariants, *Numer. Math.* **103**, 575–590 (2006).

[9] P. Chartier, A. Murua, Preserving first integrals and volume forms of additively split systems, *IMA J. Numer. Anal.* **27**, 381–405 (2007).

[10] P. Chartier, A. Murua, J.M. Sanz-Serna, Higher-Order averaging, formal series and numerical integration I: B-series, *Found. Comput. Math.* **10**, 695–727 (2010).

[11] P. Chartier, A. Murua, J.M. Sanz-Serna, A formal series approach to averaging: exponentially small error estimates, submitted.

[12] E. Hairer and Ch. Lubich, Long-time energy conservation of numerical methods for oscillatory differential equations, *SIAM J. Numer. Anal.* **38**, 414–441 (2000).

[13] E. Hairer, Ch. Lubich and G. Wanner, Geometric Numerical Integration, 2nd ed., Springer, Berlin, 2006.

[14] E. Hairer, S. P. Nørsett and G. Wanner, Solving Ordinary Differential Equations I, Nonstiff Problems, 2nd ed., Springer, Berlin, 1993.

[15] E. Hairer and G. Wanner, On the Butcher group and general multi-value methods, *Computing* **13**, 287–303 (1974).

[16] N. Jacobson, Lie Algebras, Dover, New York, 1979.

[17] P. Lochak, C. Meunier, Multiphase Averaging for Classical Systems, with Applications to Adiabatic Theorems, Springer, New York, 1988.

[18] A. Murua, Formal series and numerical integrators, part I: systems of ODEs and symplectic integrators, *Appl. Numer. Math.* **29**, 221–251 (1999).

[19] A. Murua, The Hopf algebra of rooted trees, free Lie algebras and Lie series, *Found. Comput. Math.* **6**, 387–426 (2006).

[20] A.I. Neishtadt, The separation of motions in systems with rapidly rotating phase, *J. Appl. Math. Mech.* **48**, 133–139 (1984).

[21] L.M. Perko, Higher order averaging and related methods for perturbed periodic and quasi-periodic systems, *SIAM J. Appl. Math.* **17**, 698–724 (1968).

[22] J. A. Sanders, F. Verhulst and J. Murdock, Averaging Methods in Nonlinear Dynamical Systems (2nd. ed.), Springer, New York, 2007.

[23] J. M. Sanz-Serna and M. P. Calvo, Numerical Hamiltonian Problems, Chapman and Hall, London, 1994.

[24] C. Simó, Averaging under fast quasi-periodic forcing. In: I. Seimenis (ed.) Proceedings of the NATO-ARW Integrable and Chaotic Behaviour in Hamiltonian Systems, Torun, Poland, 1993, pp. 13–34. Plenum, New York, 1994.