Averaging and computing normal forms with word series algorithms

A. Murua and J.M. Sanz-Serna

Abstract In the first part of the present work we consider periodically or quasiperiodically forced systems of the form $(d/dt)x = \varepsilon f(x, t\omega)$, where $\varepsilon \ll 1$, $\omega \in \mathbb{R}^d$ is a nonresonant vector of frequencies and $f(x, \theta)$ is 2π-periodic in each of the d components of θ (i.e. $\theta \in \mathbb{T}^d$). We describe in detail a technique for explicitly finding a change of variables $x = u(X, \theta; \varepsilon)$ and an (autonomous) averaged system $(d/dt)X = \varepsilon F(X; \varepsilon)$ so that, formally, the solutions of the given system may be expressed in terms of the solutions of the averaged system by means of the relation $x(t) = u(X(t), t\omega; \varepsilon)$. Here u and F are found as series whose terms consist of vector-valued maps weighted by suitable scalar coefficients. The maps are easily written down by combining the Fourier coefficients of f and the coefficients are found with the help of simple recursions. Furthermore these coefficients are *universal* in the sense that they do not depend on the particular f under consideration. In the second part of the contribution, we study problems of the form $(d/dt)x = g(x) + f(x)$, where one knows how to integrate the 'unperturbed' problem $(d/dt)x = g(x)$ and f is a perturbation satisfying appropriate hypotheses. It is shown how to explicitly rewrite the system in the 'normal form' $(d/dt)x = \bar{g}(x) + \bar{f}(x)$, where $\bar{g}$ and $\bar{f}$ are *commuting* vector fields and the flow of $(d/dt)x = \bar{g}(x)$ is conjugate to that of the unperturbed $(d/dt)x = g(x)$. In Hamiltonian problems the normal form directly leads to the explicit construction of formal invariants of motion. Again, $\bar{g}$, $\bar{f}$ and the invariants are written as series consisting of known vector-valued maps and universal scalar coefficients that may be found recursively.

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

J.M. Sanz-Serna
Departamento de Matemáticas, Universidad Carlos III de Madrid, E–28911 Leganés (Madrid), Spain. e-mail: jmsanzserna@gmail.com

1 Introduction

In this article we illustrate how to use word series to manipulate systems of differential equations. Specifically we deal with the questions of high-order averaging of periodically or quasiperiodically forced systems and reduction to normal form of perturbations of integrable systems. The manipulations require operations with complex numbers rather than with vector fields.

Word series are patterned after B-series [10], a well-known tool to analyse numerical integrators (see [18] for a summary of the uses of formal series in the numerical analysis of differential equations). While B-series are parameterized by rooted trees, word series [14] possess one term for each word w that may be composed with the letters of a suitable alphabet A [15]. Each term $\delta_w f_w$ of a word series is the product of a scalar coefficient δ_w and a vector field f_w. The vector fields f_w may be immediately constructed and depend on the differential system under consideration. The coefficients δ_w are *universal*, in the sense that they do not change with the particular differential system being studied. Series of *differential operators* parameterized by words (Chen-Fliess series) are very common, e.g. in control theory [9] and dynamical systems [7] and have also been used in numerical analysis (see [11] among others). As discussed in [14], word series are mathematically equivalent to Chen-Fliess series, but being series of functions they are handled in a way very similar to the way numerical analysts handle B-series. In the present work, as in [3], [4], [5], [13], the formal series techniques originally introduced to analyze numerical integrators are applied to the study of dynamical systems.

The structure of this article is as follows. The use of word series is briefly reviewed in Section 2. Section 3 addresses the problem of averaging periodically or quasiperiodically forced systems. We find a change of variables that formally reduces the system to time-independent (averaged) form. Both the change of variables and the averaged system are expressed by means of word series with universal coefficients that may be computed by means of simple recursions. The averaged system obtained in this way has favourable geometric properties. It is equivariant with respect to arbitrary changes of variables, i.e., the operations of changing variables and averaging commute. In addition averaging a Hamiltonian/divergence free/... system results in a system that is also Hamiltonian/divergence free/... Sections 4 and 5 are devoted to the reduction to normal form of general classes of perturbed problems.

Let us discuss the relation between this article and our earlier contributions. The problems envisaged here have been considered in [4]. However the treatment in [4] makes heavy use of B-series; word series results are derived, by means of the Hopf algebra techniques of [11], as a byproduct of B-series results. Here the circuitous derivations of [4] are avoided by working throughout with word series, without any reference to B-series. One of our aims when writing this article has been to provide potential users of word series techniques in application problems with a more focused, brief and clear guide than [4] provides. In addition, the class of perturbed problems considered in Sections 4 and 5 below is much wider than that considered in [4]. In [13] we have recently addressed the reduction of perturbed problems to normal forms. The treatment in [13] is based on the application of successive changes

of variables; here the normal form is directly obtained in the originally given variables. An application of word series techniques to stochastic problems is provided in [1]. The article [12] presents an application of the high-order averaging described here to a problem arising in vibrational resonance.

All the developments in the article use formal series of smooth maps. To streamline the presentation the words 'formal' and 'smooth' are often omitted. By truncating the formal expansions obtained in this article it is possible to obtain nonformal results, as in [5] or [6], but we shall not be concerned with such a task.

2 Word series

We begin by presenting the most important rules for handling word series. For proofs and additional properties of word series, the reader is referred to [14].

2.1 Defining word series

Assume that A is a finite or infinite countable set of indices (the alphabet) and that for each element (letter) $\ell \in A$, $f_\ell(y)$ is a map $f_\ell : \mathbb{C}^d \to \mathbb{C}^d$. Associated with each nonempty word $\ell_1 \cdots \ell_n$ constructed with letters from the alphabet, there is a *word basis function*. These are defined recursively by

$$f_{\ell_1 \cdots \ell_n}(y) = f'_{\ell_2 \cdots \ell_n}(y) f_{\ell_1}(y), \quad n > 1,$$

where $f'_{\ell_2 \cdots \ell_n}(y)$ is the Jacobian matrix of $f_{\ell_2 \cdots \ell_n}(y)$. For the empty word, the corresponding basis function is the identity map $y \mapsto y$. The set of all words (including the empty word $\emptyset$) will be denoted by $\mathscr{W}$ and the symbol $\mathbb{C}^{\mathscr{W}}$ will be used to refer to the vector space of all mappings $\delta : \mathscr{W} \to \mathbb{C}$. For $\delta \in \mathbb{C}^{\mathscr{W}}$ and $w \in \mathscr{W}$, δ_w is the complex number that δ associates with w. To each $\delta \in \mathbb{C}^{\mathscr{W}}$ there corresponds a *word series* (relative to the mappings f_ℓ); this is the formal series

$$W_\delta(y) = \sum_{\delta \in \mathscr{W}} \delta_w f_w(y).$$

The numbers δ_w, $w \in \mathscr{W}$, are the *coefficients* of the series.

Let us present an example. If for each letter $\ell \in A$, $\lambda_\ell(t)$ is a scalar-valued function of the real variable t, the solution of initial value problem

$$\frac{d}{dt} y = \sum_{\ell \in A} \lambda_\ell(t) f_\ell(y), \quad y(t_0) = y_0 \in \mathbb{C}^D \tag{1}$$

has a formal expansion in terms of word series given by

$$y(t) = W_{\alpha(t;t_0)}(y_0), \tag{2}$$

where, for each t, t_0, the coefficients $\alpha_w(t;t_0)$ are the iterated integrals

$$\alpha_{\ell_1 \cdots \ell_n}(t;t_0) = \int_{t_0}^{t} dt_n\, \lambda_{\ell_n}(t_n) \int_{t_0}^{t_n} dt_{n-1}\, \lambda_{\ell_{n-1}}(t_{n-1}) \cdots \int_{t_0}^{t_2} dt_1\, \lambda_{\ell_1}(t_1). \tag{3}$$

This series representation, whose standard derivation may be seen in e.g. [5] or [14], is essentially the Chen series used in control theory. (An alternative derivation is presented below.) Of much importance in what follows is the fact that the coefficients $\alpha_w(t;t_0)$ depend only on the $\lambda_\ell(t)$ in (1) and do not change with the vector fields $f_\ell(y)$; on the contrary, the word basis functions $f_w(y)$ depend on the $f_\ell(y)$ and do not change with the $\lambda_\ell(t)$.

2.2 The convolution product

The *convolution product* $\delta \star \delta' \in \mathbb{C}^{\mathscr{W}}$ of two elements $\delta, \delta' \in \mathbb{C}^{\mathscr{W}}$ is defined by

$$(\delta \star \delta')_{\ell_1 \cdots \ell_n} = \delta_\emptyset \delta'_{\ell_1 \cdots \ell_n} + \sum_{j=1}^{n-1} \delta_{\ell_1 \cdots \ell_j} \delta'_{\ell_{j+1} \cdots \ell_n} + \delta_{\ell_1 \cdots \ell_n} \delta'_\emptyset, \quad n \geq 1$$

$((\delta \star \delta')_\emptyset = \delta_\emptyset \delta'_\emptyset)$. The operation $\star$ is not commutative, but it is associative and has a unit (the element $\mathbb{1} \in \mathbb{C}^{\mathscr{W}}$ with $\mathbb{1}_\emptyset = 1$ and $\mathbb{1}_w = 0$ for $w \neq \emptyset$).

If w and w' are words, their *shuffle product* will be denoted by $w \sqcup\!\sqcup w'$; this is the formal sum of all words that may be formed by interleaving the letters of w with those of w' without altering the order in which those letters appear within w or w' (e.g., $\ell m \sqcup\!\sqcup n = \ell m n + \ell n m + n \ell m$). The set $\mathscr{G}$ consists of those $\gamma \in \mathbb{C}^{\mathscr{W}}$ that satisfy the following *shuffle relations:* $\gamma_\emptyset = 1$ and, for each $w, w' \in \mathscr{W}$,

$$\gamma_w \gamma_{w'} = \sum_{j=1}^{N} \gamma_{w_j} \qquad \text{if} \qquad w \sqcup\!\sqcup w' = \sum_{j=1}^{N} w_j.$$

This set is a group for the operation $\star$. For each t and t_0 the element $\alpha(t;t_0) \in \mathbb{C}^{\mathscr{W}}$ in (3) belongs to the group $\mathscr{G}$.

For $\gamma \in \mathscr{G}$, $\delta \in \mathbb{C}^{\mathscr{W}}$,

$$W_\delta\big(W_\gamma(x)\big) = W_{\gamma \star \delta}(x). \tag{4}$$

In words: the substitution of $W_\gamma(x)$ in an arbitrary word series $W_\delta(x)$ gives rise to a new word series whose coefficients are given by the convolution product $\gamma \star \delta$. We emphasize that this result does not hold for arbitrary $\gamma \in \mathbb{C}^{\mathscr{W}}$, the hypothesis $\gamma \in \mathscr{G}$ is essential.

Another property of the word series $W_\gamma(y)$ with $\gamma \in \mathscr{G}$ is its *equivariance* [18] with respect to arbitrary changes of variables $y = C(\bar{y})$. If $\bar{f}_\ell(\bar{y})$ is the result (pullback) of changing variables in the field $f_\ell(y)$, i.e.,

$$\bar{f}_\ell(\bar{y}) = C'(\bar{y})^{-1} f_\ell(C(\bar{y})),$$

and $\bar{W}_\gamma(\bar{y})$ denotes the word series with coefficients γ_w constructed from the fields $\bar{f}_\ell(\bar{y})$, then

$$C\big(\bar{W}_\gamma(\bar{y})\big) = W_\gamma(C(\bar{y})).$$

We denote by $\mathfrak{g}$ the vector subspace of $\mathbb{C}^{\mathcal{W}}$ consisting of those β that satisfy the following shuffle relations: $\beta_\emptyset = 0$ and for each pair of nonempty words w, w',

$$\sum_{j=1}^{N} \beta_{w_j} = 0 \qquad \text{if} \qquad w \,\text{⧢}\, w' = \sum_{j=1}^{N} w_j.$$

It is easily proved that the elements $\beta \in \mathfrak{g}$ are precisely the velocities $(d/dt)\gamma(0)$ at $t = 0$ of the smooth curves $t \mapsto \gamma(t) \in \mathscr{G}$ with $\gamma(0) = \mathbb{1}$, i.e., if $\mathscr{G}$ is formally viewed as a Lie group, then $\mathfrak{g}$ is the corresponding Lie algebra. In fact, $\mathscr{G}$ and $\mathfrak{g}$ are the group of characters and the Lie algebra of infinitesimal characters of the shuffle Hopf algebra (see [14] for details).

2.3 Universal formulations

Let us consider once more the differential system (1). Define, for fixed t, $\beta(t) \in \mathfrak{g} \subset \mathbb{C}^{\mathcal{W}}$ by $\beta_\ell(t) = \lambda_\ell(t)$, for each $\ell \in A$, and $\beta_w(t) = 0$ if the word w is empty or has ≥ 2 letters. Then the right hand-side of (1) is simply the word series $W_{\beta(t)}(y)$. We look for the solution $y(t)$ in the word series form (2), with undetermined coefficients $\alpha_w(t;t_0)$ that have to be determined and *have to belong* to the group $\mathscr{G}$. By using the formula (4), we may write

$$\frac{\partial}{\partial t} W_{\alpha(t;t_0)}(y_0) = W_{\beta(t)}\big(W_{\alpha(t;t_0)}(y_0)\big) = W_{\alpha(t;t_0)\star\beta(t)}(y_0),$$
$$W_{\alpha(t_0,t_0)}(y_0) = y_0 = W_{\mathbb{1}}(y_0),$$

and these equations will be satisfied if

$$\frac{\partial}{\partial t}\alpha(t;t_0) = \alpha(t;t_0) \star \beta(t), \qquad \alpha(t_0,t_0) = \mathbb{1}. \tag{5}$$

This is an initial value problem for the curve $t \mapsto \alpha(t;t_0)$ in the group $\mathscr{G}$ (t_0 is a parameter) and may be uniquely solved by successively determining the values $\alpha_w(t;t_0)$ for words of increasing length. In fact, for the empty word, the requirement $\alpha(t;t_0) \in \mathscr{G}$ implies $\alpha_\emptyset(t;t_0) = 1$. For words with one letter $\ell \in A$, using $\beta_\emptyset(t) = 0$ and the definition of the convolution product $\star$, we have the conditions

$$\frac{\partial}{\partial t}\alpha_\ell(t;t_0) = \beta_\ell(t)\alpha_\emptyset(t;t_0) = \lambda_\ell(t), \qquad \alpha_\ell(t_0,t_0) = 0,$$

that lead to

$$\alpha_\ell(t;t_0) = \int_{t_0}^t dt_1 \lambda_\ell(t_1).$$

This procedure may be continued (see [14] for details) to determine uniquely $\alpha_w(t;t_0)$ for all words $w \in \mathscr{W}$. In addition, for each t and t_0, the element $\alpha(t;t_0) \in \mathbb{C}^{\mathscr{W}}$ found in this way belongs to $\mathscr{G}$, as it was desired. Of course this element coincides with that defined in (3).

In going from (1) to (5) we move from an initial value problem for the vector-valued function $y(t)$ to a seemingly more complicated initial value problem for the function $\alpha(t)$ with values in $\mathscr{G}$. However the abstract problem in $\mathscr{G}$ is linear and easily solvable. Of equal importance to us is the fact that (5) is *universal* in the sense that, once it has been integrated, one readily writes, by changing the word basis functions, the solution (2) of *each* problem obtained by replacing in (1) the mappings $f_\ell(y)$ by other choices. In particular the universal character of the formulation implies that (5) is independent of the dimension D of (1).

3 Averaging of quasiperiodically forced systems

In this section, we consider the oscillatory initial value problem

$$\frac{d}{dt}y = \varepsilon f(y, t\omega), \qquad y(t_0) = y_0 \in \mathbb{C}^D, \tag{6}$$

in a long interval $t_0 \le t \le t_0 + L/\varepsilon$. The vector field $f(y, \theta)$ is 2π-periodic in each of the scalar components (angles) $\theta_1, \ldots, \theta_d$ of θ (i.e., $\theta \in \mathbb{T}^d$) and ω is a constant vector of frequencies $\omega_1, \ldots, \omega_d$. These are assumed to be *nonresonant* i.e. $\mathbf{k} \cdot \omega \ne 0$ for each multiindex $\mathbf{k} \in \mathbb{Z}^d$, $\mathbf{k} \ne \mathbf{0}$; resonant problems may be rewritten in nonresonant form by reducing the number of frequencies. Thus the forcing in (6) is quasiperiodic if $d > 1$ and periodic if $d = 1$. Our aim is to find a time-dependent change of variables $y = U(Y, t\omega; \varepsilon)$ that formally brings the differential system (6) into autonomous form [16]. Our approach is based on a *universal* formulation, analogous to the one we used above to deal with (1).

3.1 The solution of the oscillatory problem

After Fourier expanding

$$f(y, \theta) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \exp(i\mathbf{k} \cdot \theta)\widehat{f}_{\mathbf{k}}(y),$$

the problem (6) becomes a particular case of (1); each letter ℓ is a multiindex $\mathbf{k} \in \mathbb{Z}^d$, $f_\ell(y) = f_{\mathbf{k}}(y) = \varepsilon \widehat{f}_{\mathbf{k}}(y)$, and

$$\lambda_\ell(t) = \exp(i\mathbf{k}\cdot\omega t). \tag{7}$$

Each word basis function $f_w(y)$ contains the factor ε^n if w has length n. The first few coefficients (iterated integrals) $\alpha_w(t;t_0)$ in (3) are easily computed; here are a few instances

$$
\begin{aligned}
\alpha_\emptyset(t;t_0) &= 1, \\
\alpha_{\mathbf{0}}(t;t_0) &= t - t_0, \\
\alpha_{\mathbf{k}}(t;t_0) &= \frac{i\big(\exp(i\mathbf{k}\cdot\omega t_0) - \exp(i\mathbf{k}\cdot\omega t)\big)}{\mathbf{k}\cdot\omega}, \quad \mathbf{k} \neq \mathbf{0}, \\
\alpha_{\mathbf{00}}(t;t_0) &= \frac{(t-t_0)^2}{2}, \\
\alpha_{\mathbf{kl}}(t;t_0) &= \frac{i(t-t_0)}{\mathbf{k}\cdot\omega} + \frac{1 - \exp(i\mathbf{k}\cdot\omega t)\exp(-i\mathbf{k}\cdot\omega t_0)}{(\mathbf{k}\cdot\omega)^2}, \quad \mathbf{k} \neq \mathbf{0}, \mathbf{l} = -\mathbf{k}. \tag{8}
\end{aligned}
$$

Note the oscillatory components present in some of the coefficients.

The following result shows how the coefficients $\alpha_w(t;t_0)$ can be determined recursively without explicitly carrying out the integrations in (3).

Proposition 1. *The coefficients (3) with the $\lambda_\ell(t)$ given by (7) are uniquely determined by the recursive formulas*

$$
\begin{aligned}
\alpha_{\mathbf{k}}(t;t_0) &= \frac{i\big(\exp(i\mathbf{k}\cdot\omega t_0) - \exp(i\mathbf{k}\cdot\omega t)\big)}{\mathbf{k}\cdot\omega}, \\
\alpha_{\mathbf{0}^r}(t;t_0) &= (t-t_0)^r/r!, \\
\alpha_{\mathbf{0}^r\mathbf{k}}(t;t_0) &= \frac{i}{\mathbf{k}\cdot\omega}\big(\alpha_{\mathbf{0}^{r-1}\mathbf{k}}(t;t_0) - \alpha_{\mathbf{0}^r}(t;t_0)e^{i\mathbf{k}\cdot\omega t}\big), \\
\alpha_{\mathbf{kl}_1\cdots\mathbf{l}_s}(t;t_0) &= \frac{i}{\mathbf{k}\cdot\omega}\big(e^{i\mathbf{k}\cdot\omega t_0}\alpha_{\mathbf{l}_1\cdots\mathbf{l}_s}(t;t_0) - \alpha_{(\mathbf{k}+\mathbf{l}_1)\mathbf{l}_2\cdots\mathbf{l}_s}(t;t_0)\big), \\
\alpha_{\mathbf{0}^r\mathbf{kl}_1\cdots\mathbf{l}_s}(t;t_0) &= \frac{i}{\mathbf{k}\cdot\omega}\big(\alpha_{\mathbf{0}^{r-1}\mathbf{kl}_1\cdots\mathbf{l}_s}(t;t_0) - \alpha_{\mathbf{0}^r(\mathbf{k}+\mathbf{l}_1)\mathbf{l}_2\cdots\mathbf{l}_s}(t;t_0)\big),
\end{aligned} \tag{9}
$$

where $r \geq 1$, $\mathbf{k} \in \mathbb{Z}^d\backslash\{\mathbf{0}\}$, and $\mathbf{l}_1,\dots,\mathbf{l}_s \in \mathbb{Z}^d$.

Proof. It is useful to point out that the formulas (9) are found by evaluating the innermost integral in (3). To prove the proposition we show that the coefficients $\alpha_{\mathbf{k}_1\cdots\mathbf{k}_n}(t;t_0)$ uniquely determined by (9) coincide with those in (3). The latter satisfy, for all words $w = \mathbf{k}_1\cdots\mathbf{k}_n$,

$$\frac{d}{dt}\alpha_{\mathbf{k}_1\cdots\mathbf{k}_n}(t;t_0) = \exp(i\mathbf{k}_n\cdot\omega t)\alpha_{\mathbf{k}_1\cdots\mathbf{k}_{n-1}}(t;t_0), \qquad \alpha_{\mathbf{k}_1\cdots\mathbf{k}_n}(t_0;t_0) = 0. \tag{10}$$

We prove by induction on n that the coefficients in (9) also satisfy (10). One can trivially check the case $n = 1$. For each word $w = \mathbf{k}_1\cdots\mathbf{k}_n$ with $n > 1$, one arrives at (10) by differentiating with respect to t both sides of the equality in (9) that determines $\alpha_{\mathbf{k}_1\cdots\mathbf{k}_n}(t;t_0)$ and applying the induction hypothesis. $\square$

3.2 The transport equation

It follows from Proposition 1 that each $\alpha_w(t;t_0)$ is of the form

$$\alpha_w(t;t_0) = \Gamma_w(t-t_0, \omega t; \omega t_0), \tag{11}$$

where $\Gamma_w(\tau, \theta; \theta_0)$ is a suitable scalar-valued function, which is, as a function of $\tau \in \mathbb{R}$, a polynomial and as a function of $\theta \in \mathbb{T}^d$ (or of $\theta_0 \in \mathbb{T}^d$) a trigonometric polynomial. For instance, for $\mathbf{k} \neq \mathbf{0}, \mathbf{l} = -\mathbf{k}$, (see (8)),

$$\Gamma_{\mathbf{kl}}(\tau, \theta; \theta_0) = \frac{i\tau}{\mathbf{k}\cdot\omega} + \frac{1 - \exp(i\mathbf{k}\cdot\theta)\exp(-i\mathbf{k}\cdot\theta_0)}{(\mathbf{k}\cdot\omega)^2}.$$

Of course the Γ_w can be found recursively by mimicking (9). The following result summarizes this discussion.

Theorem 1. *Define, for each $w \in \mathscr{W}$, $\Gamma_w(\tau, \theta; \theta_0)$ by means of the following recursions. $\Gamma_\emptyset(\tau, \theta; \theta_0) = 1$, and given $r \geq 1$, $\mathbf{k} \in \mathbb{Z}^d - \{\mathbf{0}\}$, and $\mathbf{l}_1, \ldots, \mathbf{l}_s \in \mathbb{Z}^d$,*

$$\Gamma_{\mathbf{k}}(\tau, \theta; \theta_0) = \frac{i}{\mathbf{k}\cdot\omega}(e^{i\mathbf{k}\cdot\theta_0} - e^{i\mathbf{k}\cdot\theta}),$$

$$\Gamma_{\mathbf{0}^r}(\tau, \theta; \theta_0) = \tau^r/r!,$$

$$\Gamma_{\mathbf{0}^r\mathbf{k}}(\tau, \theta; \theta_0) = \frac{i}{\mathbf{k}\cdot\omega}(\Gamma_{\mathbf{0}^{r-1}\mathbf{k}}(\tau, \theta; \theta_0) - \Gamma_{\mathbf{0}^r}(\tau, \theta; \theta_0)e^{i\mathbf{k}\cdot\theta}), \tag{12}$$

$$\Gamma_{\mathbf{kl}_1\cdots\mathbf{l}_s}(\tau, \theta; \theta_0) = \frac{i}{\mathbf{k}\cdot\omega}(e^{i\mathbf{k}\cdot\theta_0}\Gamma_{\mathbf{l}_1\cdots\mathbf{l}_s}(\tau, \theta; \theta_0) - \Gamma_{(\mathbf{k}+\mathbf{l}_1)\mathbf{l}_2\cdots\mathbf{l}_s}(\tau, \theta; \theta_0)),$$

$$\Gamma_{\mathbf{0}^r\mathbf{kl}_1\cdots\mathbf{l}_s}(\tau, \theta; \theta_0) = \frac{i}{\mathbf{k}\cdot\omega}(\Gamma_{\mathbf{0}^{r-1}\mathbf{kl}_1\cdots\mathbf{l}_s}(\tau, \theta; \theta_0) - \Gamma_{\mathbf{0}^r(\mathbf{k}+\mathbf{l}_1)\mathbf{l}_2\cdots\mathbf{l}_s}(\tau, \theta; \theta_0)).$$

Then, for each $w \in \mathscr{W}$, $\Gamma_w(\tau, \theta; \theta_0)$ is a polynomial in τ and a trigonometric polynomial in θ and in θ_0 and the coefficient $\alpha_w(t;t_0)$ of the oscillatory solution satisfies (11).

Substituting (11) in the initial value problem (5) that characterizes $\alpha(t;t_0)$, we find, after using the chain rule,

$$\frac{\partial}{\partial\tau}\Gamma(t-t_0, t\omega; t_0\omega) + \omega\cdot\nabla_\theta\Gamma(t-t_0, t\omega; t_0\omega) = \Gamma(t-t_0, t\omega; t_0\omega) \star B(t\omega),$$
$$\Gamma(0, t_0\omega; t_0\omega) = \mathbb{1},$$

where $B(\theta) \in \mathfrak{g}$ is defined as $B_{\mathbf{k}}(\theta) = \exp(i\mathbf{k}\cdot\theta)$, $\mathbf{k} \in \mathbb{Z}^d$, and $B_w(\theta) = 0$ if the length of w is not 1. We thus have that for all $(\tau, \theta; \theta_0)$ of the form $(t-t_0, t\omega; t_0\omega)$, the following equation is valid:

$$\frac{\partial}{\partial\tau}\Gamma(\tau, \theta; \theta_0) + \omega\cdot\nabla_\theta\Gamma(\tau, \theta; \theta_0) = \Gamma(\tau, \theta; \theta_0) \star B(\theta), \quad \Gamma(0, \theta_0; \theta_0) = \mathbb{1}. \tag{13}$$

Actually, it can be proved, by mimicking the proof of Proposition 1 (with d/dt replaced by the operator $\partial/\partial\tau + \omega\cdot\nabla_\theta$), that (13) holds for arbitrary $(\tau,\theta;\theta_0) \in \mathbb{R}\times\mathbb{T}^d\times\mathbb{T}^d$.

We have thus found a *transport equation* for Γ as a function of τ and θ (θ_0 plays the role of a parameter).[1] For this partial differential equation, a standard initial condition would prescribe the value of $\Gamma(0,\theta;\theta_0)$ as a function of $\theta \in \mathbb{T}^d$ (and of the parameter θ_0); in (13), $\Gamma(0,\theta;\theta_0)$ is only given at the single point $\theta = \theta_0$. Therefore (13) may be expected to have many solutions; only one of them is such that, for each $w \in \mathcal{W}$, $\Gamma_w(\tau,\theta;\theta_0)$ depends polynomially on τ as we shall establish in Proposition 2. We shall use an auxiliary result whose simple proof will be ommitted (cf. Lemma 2.4 in [4]):

Lemma 1. *Let the vector $\omega \in \mathbb{R}^d$ be nonresonant. If a smooth function $z : \mathbb{R}\times\mathbb{T}^d \to \mathbb{C}$ satisfies*

$$\frac{\partial}{\partial\tau}z(\tau,\theta) + \omega\cdot\nabla_\theta z(\tau,\theta) = 0, \quad z(0,\theta_0) = 0,$$

and $z(\tau,\theta)$ is polynomial in τ, then $z(\tau,\theta)$ is identically zero.

Proposition 2. *The function $\Gamma(\tau,\theta;\theta_0)$ given in Theorem 1 is the unique solution of problem (13) such that each $\Gamma_w(\tau,\theta;\theta_0))$, $w \in \mathcal{W}$, is smooth in θ and polynomial in τ.*

Proof. Let $\delta(\tau,\theta;\theta_0)$ denote the difference of two solutions of (13). Then, for each $w \in \mathcal{W}$, $\delta_w(0,\theta_0;\theta_0) = 0$ and

$$\frac{\partial}{\partial\tau}\delta_w(\tau,\theta;\theta_0) + \omega\cdot\nabla_\theta\delta_w(\tau,\theta;\theta_0)$$

vanishes provided that the value of $\delta(\tau,\theta;\theta_0)$ at words with less letters than w vanish identically. Lemma 1 then allows us to prove, by induction on the number of letters, that $\delta_w(\tau,\theta;\theta_0) \equiv 0$ for all $w \in \mathcal{W}$. See [4], Section 2.4 for a similar proof. $\square$

The transport problem is used in the proof of the following two theorems, which in turn play an important role in averaging.

Theorem 2. *For each $\tau \in \mathbb{R}$, $\theta \in \mathbb{T}^d$, $\theta_0 \in \mathbb{T}^d$, the element $\Gamma(\tau,\theta;\theta_0) \in \mathbb{C}^{\mathcal{W}}$ belongs to $\mathcal{G}$.*

Proof. The proof is very similar to the proof given in Section 6.1.4 of [14] for nonautonomous ordinary linear differential equations in $\mathcal{G}$. We have to prove that

$$\sum_j \Gamma_{w_j}(\tau,\theta;\theta_0) = \Gamma_w(\tau,\theta;\theta_0)\Gamma_{w'}(\tau,\theta;\theta_0) \tag{14}$$

---

[1] The presence of this parameter is linked to the fact that the transport equation is nonautonomous in the variable θ.

for $w, w' \in \mathscr{W}$, with $w \sqcup w' = \sum_j w_j$. This is established by induction on the sum of the number of letters of w and w'. Proceeding as in [14], by application of the induction hypothesis one arrives at

$$(\frac{\partial}{\partial \tau} + \omega \cdot \nabla_\theta) \left(\sum_j \Gamma_{w_j}(\tau, \theta; \theta_0) - \Gamma_w(\tau, \theta; \theta_0) \Gamma_{w'}(\tau, \theta; \theta_0) \right) = 0.$$

Since (14) holds at $(\tau, \theta) = (0, \theta_0)$, Lemma 1 implies that it does so for each value of $(\tau, \theta) \in \mathbb{R} \times \mathbb{T}^d$. $\square$

Theorem 3. *For arbitrary $\tau_1, \tau_2 \in \mathbb{R}$ and $\theta_0, \theta_1, \theta_2 \in \mathbb{T}^d$,*

$$\Gamma(\tau_1, \theta_1; \theta_0) \star \Gamma(\tau_2, \theta_2; \theta_1) = \Gamma(\tau_1 + \tau_2, \theta_2; \theta_0).$$

Proof. One can check that both $\gamma(\tau, \theta) = \Gamma(\tau_1, \theta_1; \theta_0) \star \Gamma(\tau - \tau_1, \theta; \theta_1)$ and $\gamma(\tau, \theta) = \Gamma(\tau, \theta; \theta_0)$ satisfy the following three conditions:

- for each $w \in \mathscr{W}$, $\gamma_w(\tau, \theta)$ is smooth in θ and depends polynomially on τ,
- $\gamma(\tau_1, \theta_1) = \Gamma(\tau_1, \theta_1; \theta_0)$, and
- for all $\tau \in \mathbb{R}$ and all $\theta \in \mathbb{T}^d$,

$$\frac{\partial}{\partial \tau} \gamma(\tau, \theta) + \omega \cdot \nabla_\theta \gamma(\tau, \theta) = \gamma(\tau, \theta) \star B(\theta).$$

Proceeding as in the proof of Proposition 2, one concludes that there is a unique $\gamma(\tau, \theta)$ satisfying the three conditions above. The required result is thus obtained by setting $\tau = \tau_1 + \tau_2$ and $\theta = \theta_2$. $\square$

In particular

$$\Gamma(\tau_1, \theta_0; \theta_0) \star \Gamma(\tau_2, \theta_0; \theta_0) = \Gamma(\tau_1 + \tau_2, \theta_0; \theta_0),$$
$$\Gamma(0, \theta_1; \theta_0) \star \Gamma(0, \theta_2; \theta_1) = \Gamma(0, \theta_2; \theta_0),$$

for arbitrary $\tau_1, \tau_2 \in \mathbb{R}$, $\theta_0, \theta_1, \theta_2 \in \mathbb{T}^d$. The first of these identities shows that, as τ varies with θ_0 fixed, the elements $\Gamma(\tau, \theta_0; \theta_0)$ form a one-parameter subgroup of $\mathscr{G}$. The second identity is similar to what is sometimes called two-parameter group property of the solution operator of nonautonomous differential equations. Of course these identities reflect the fact that the transport equation is autonomous in τ and nonautonomous in θ.

3.3 The averaged system and the change of variables

Before we average the oscillatory problem (6), we shall do so with the corresponding universal problem (5) in $\mathscr{G}$, whose solution α has been represented in (11) by means of the auxiliary function $\Gamma(\tau, \theta; \theta_0)$. Note that the oscillatory nature of α is caused

by the second argument in Γ (each Γ_w, $w \in \mathscr{W}$, is a polynomial τ). Consider then the $\mathscr{G}$-valued function of t defined by

$$\bar{\alpha}(t;t_0) = \Gamma(t - t_0, t_0\omega; t_0\omega), \tag{15}$$

where the second argument in Γ has been frozen at its initial value. This satisfies $\bar{\alpha}(t_0;t_0) = \Gamma(0, t_0\omega; t_0\omega)$, or, from Proposition 2, $\bar{\alpha}(t_0;t_0) = \mathbb{1}$ so that $\bar{\alpha}(t;t_0)$ coincides with $\alpha(t;t_0)$ at the initial time $t = t_0$. Furthermore, due to the trigonometric dependence on θ, if $d = 1$ (periodic case), $\bar{\alpha}(t;t_0) = \Gamma(t - t_0, t_0\omega; t_0\omega)$ coincides with $\alpha(t;t_0) = \Gamma(t - t_0, t\omega; t_0\omega)$ at all times of the form $t = t_0 + 2k\pi/\omega$, k integer. If $d > 1$ (quasiperiodic case), as t varies, the point $t\omega \in \mathbb{T}^d$ never returns to the initial position $t_0\omega$; however it returns infinitely often to the neighborhood of $t_0\omega$. To sum up, the nonoscillatory $\bar{\alpha}(t;t_0)$ is a good description of the long-term evolution of $\alpha(t;t_0)$ and, in fact, we shall presently arrange things in such a way that $\bar{\alpha}(t;t_0)$ is the solution of the averaged version of the problem (5).

Having identified the averaged *solution,* let us find the averaged *problem.* From Theorem 3, for each τ_1 and τ_2,

$$\bar{\alpha}(t_0 + \tau_1 + \tau_2; t_0) = \bar{\alpha}(t_0 + \tau_1; t_0) \star \bar{\alpha}(t_0 + \tau_2; t_0)$$

so that, as τ varies, the elements $\bar{\alpha}(t_0 + \tau; t_0)$ form a (commutative) one-parameter group $\subset \mathscr{G}$. Therefore $\bar{\alpha}(t;t_0)$ is the solution of the *autonomous problem*

$$\frac{d}{dt}\bar{\alpha}(t;t_0) = \bar{\alpha}(t;t_0) \star \bar{\beta}(t_0), \qquad \bar{\alpha}(t_0;t_0) = \mathbb{1}, \tag{16}$$

with

$$\bar{\beta}(t_0) = \frac{d}{dt}\bar{\alpha}(t;t_0)\bigg|_{t=t_0}. \tag{17}$$

For completeness we include here a proof of this fact, which is well known in the theory of differential equations,

$$\frac{d}{dt'}\bar{\alpha}(t';t_0) = \frac{d}{dt}\bar{\alpha}(t' + t - t_0; t_0)\bigg|_{t=t_0} = \frac{d}{dt}\bar{\alpha}(t';t_0) \star \bar{\alpha}(t;t_0)\bigg|_{t=t_0} = \bar{\alpha}(t';t_0) \star \bar{\beta}(t_0).$$

Note that (17) implies that $\bar{\beta}(t_0) \in \mathfrak{g}$.

After having found the averaged problem (16), we invoke once more Theorem 3 and write

$$\begin{aligned}
\alpha(t;t_0) &= \Gamma(t - t_0, t\omega; t_0\omega) \\
&= \Gamma(t - t_0, t_0\omega; t_0\omega) \star \Gamma(0, t\omega; t_0\omega) \\
&= \bar{\alpha}(t;t_0) \star \Gamma(0, t\omega; t_0\omega).
\end{aligned}$$

Thus, if we define

$$\kappa(\theta; t_0) = \Gamma(0, \theta; t_0\omega), \tag{18}$$

then κ depends periodically on the components of θ and $\kappa(t\omega;t_0)$ relates the averaged solution $\bar{\alpha}$ and the oscillatory solution α in the following way:

$$\alpha(t;t_0) = \bar{\alpha}(t;t_0) \star \kappa(t\omega;t_0). \tag{19}$$

To sum up, we have proved:

Theorem 4. *For $\theta \in \mathbb{T}^d$ define $\kappa(\theta;t_0) \in \mathscr{G}$ by (18). Then the solution of the problem (5) has the representation (19), where $\bar{\alpha}(t;t_0)$ satisfies the autonomous (averaged) initial value problem (16)–(17) with $\bar{\beta}(t_0) \in \mathfrak{g}$. Furthermore $\bar{\alpha}(t;t_0)$ may be found by means of (15).*

By inserting the word basis functions to obtain the corresponding series and recalling that the operation $\star$ for the coefficients represents the composition of the series, we conclude:

Theorem 5. *With the notation of the preceding theorem, the solution of (6) may be represented as*

$$y(t) = W_{\kappa(t\omega;t_0)}(Y(t))$$

where $Y(t) = W_{\bar{\alpha}(t;t_0)}(y_0)$ solves the autonomous (averaged) initial value problem

$$\frac{d}{dt}Y = W_{\bar{\beta}(t_0)}(Y), \qquad Y(t_0) = y_0.$$

3.4 Geometric properties

Since $\bar{\beta}(t_0)$ is in the Lie algebra $\mathfrak{g}$, the Dynkin–Specht–Wever theorem [8], implies that the word series for the averaged vector field may be rewritten in terms of iterated Lie-Jacobi brackets

$$W_{\bar{\beta}(t_0)}(y) = \sum_{r=1}^{\infty} \sum_{\mathbf{k}_1,\dots,\mathbf{k}_r \in Z^d} \frac{\varepsilon^r}{r} \beta_{\mathbf{k}_1\cdots\mathbf{k}_r}(t_0) [[\cdots[[f_{\mathbf{k}_1},f_{\mathbf{k}_2}],f_{\mathbf{k}_3}]\cdots],f_{\mathbf{k}_r}](y). \tag{20}$$

(The bracket is defined by $[f,g](y) = g'(x)f(y) - f'(y)g(y)$.)

It follows from (20) that if all the $f_{\mathbf{k}}$ belong to a given Lie subalgebra of the Lie algebra of all vector fields (e.g., if they are all Hamiltonian or all divergence free), then the averaged vector field will also lie in that subalgebra (i.e., will be Hamilonian or divergence free).

Additionally, the averaging procedure described above is *equivariant* with respect to arbitrary changes of variables: changing variables $y = C(\bar{y})$ in the oscillatory problem, followed by averaging, yields the same result as changing variables in the averaged system. This is a consequence of the equivariance of word series with coefficients in $\mathscr{G}$.

3.5 Finding the coefficients

From (15), (respectively (18)) the quantities $\bar{\alpha}_w(t;t_0)$ (respectively $\kappa_w(\theta;t_0)$), $w \in \mathscr{W}$, may be found recursively by setting $\tau = t - t_0$, $\theta = \theta_0 = t_0\omega$ (respectively $\tau = 0$, $\theta_0 = t_0\omega$) in the formulas for $\Gamma_w(\tau,\theta;\theta_0)$ provided in Theorem 1. The following recurrences for $\bar{\beta}(t_0)$ are easily found via (17).

Theorem 6. *Given* $r \geq 1$, $\mathbf{k} \in \mathbb{Z}^d \backslash \{0\}$, *and* $\mathbf{l}_1,\ldots,\mathbf{l}_s \in \mathbb{Z}^d$,

$$\bar{\beta}_{\mathbf{k}}(t_0) = 0,$$
$$\bar{\beta}_{\mathbf{0}}(t_0) = 1,$$
$$\bar{\beta}_{\mathbf{0}^{r+1}}(t_0) = 0,$$
$$\bar{\beta}_{\mathbf{0}^r\mathbf{k}}(t_0) = \frac{i}{\mathbf{k}\cdot\omega}(\bar{\beta}_{\mathbf{0}^{r-1}\mathbf{k}}(t_0) - \bar{\beta}_{\mathbf{0}^r}(t_0)e^{i\mathbf{k}\cdot\omega t_0}),$$
$$\bar{\beta}_{\mathbf{k}\mathbf{l}_1\cdots\mathbf{l}_s}(t_0) = \frac{i}{\mathbf{k}\cdot\omega}(e^{i\mathbf{k}\cdot\omega t_0}\bar{\beta}_{\mathbf{l}_1\cdots\mathbf{l}_s}(t_0) - \bar{\beta}_{(\mathbf{k}+\mathbf{l}_1)\mathbf{l}_2\cdots\mathbf{l}_s}(t_0)),$$
$$\bar{\beta}_{\mathbf{0}^r\mathbf{k}\mathbf{l}_1\cdots\mathbf{l}_s}(t_0) = \frac{i}{\mathbf{k}\cdot\omega}(\bar{\beta}_{\mathbf{0}^{r-1}\mathbf{k}\mathbf{l}_1\cdots\mathbf{l}_s}(t_0) - \bar{\beta}_{\mathbf{0}^r(\mathbf{k}+\mathbf{l}_1)\mathbf{l}_2\cdots\mathbf{l}_s}(t_0)).$$

In the particular case $t_0 = 0$, after computing the coefficients $\bar{\beta}_w(0)$ for words with ≤ 3 letters by means of the formulas in the theorem, we obtain, with the help of the Jacobi identity for the bracket and the shuffle relations, the following explicit formula for the averaged system:

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

where

$$F_2 = \sum_{\mathbf{k}>-\mathbf{k}} \frac{i}{\mathbf{k}\cdot\omega}([f_{\mathbf{k}} - f_{-\mathbf{k}}, f_{\mathbf{0}}] + [f_{-\mathbf{k}}, f_{\mathbf{k}}]),$$

$$F_3 = \sum_{\mathbf{k}\neq\mathbf{0}} \frac{1}{(\mathbf{k}\cdot\omega)^2}\left([f_{\mathbf{0}},[f_{\mathbf{0}},f_{\mathbf{k}}]] + [f_{\mathbf{k}},[f_{\mathbf{k}},f_{-\mathbf{k}}]] - \frac{1}{2}[f_{\mathbf{k}},[f_{\mathbf{k}},f_{-2\mathbf{k}}]] + [f_{-\mathbf{k}},[f_{\mathbf{k}},f_{\mathbf{0}}]]\right)$$

$$+ \sum_{\mathbf{0}\neq\mathbf{m}\neq-\mathbf{l}\neq\mathbf{0}} \frac{-1}{(\mathbf{l}\cdot\omega)((\mathbf{m}+\mathbf{l})\cdot\omega)} [f_{\mathbf{m}},[f_{\mathbf{l}},f_{\mathbf{0}}]]$$

$$+ \sum_{-\mathbf{l}>\mathbf{k}<\mathbf{l},\ \mathbf{k}\neq\mathbf{0}} \frac{1}{(\mathbf{k}\cdot\omega)(\mathbf{l}\cdot\omega)} [f_{-\mathbf{l}},[f_{\mathbf{l}},f_{\mathbf{k}}]]$$

$$+ \sum_{\substack{\mathbf{m}>\mathbf{k}<-\mathbf{k}\\ \mathbf{m}+\mathbf{k}\neq\mathbf{0}}} \frac{-1}{(\mathbf{k}\cdot\omega)(\mathbf{m}\cdot\omega)} [f_{\mathbf{m}},[f_{-\mathbf{k}},f_{\mathbf{k}}]]$$

$$+ \sum_{\substack{\mathbf{0}\neq\mathbf{m}\neq\pm\mathbf{l}\neq\mathbf{0}\\ \mathbf{m}>-\mathbf{m}-\mathbf{l}<\mathbf{l}}} \frac{-1}{(\mathbf{m}\cdot\omega)((\mathbf{m}+\mathbf{l})\cdot\omega)} [f_{\mathbf{m}},[f_{\mathbf{l}},f_{-\mathbf{m}-\mathbf{l}}]].$$

In these formulas $<$ is some total ordering in the set of multi-indices $\mathbb{Z}^d$ such that $\mathbf{k} > \mathbf{0}$ for $\mathbf{k} \neq \mathbf{0}$.

3.6 Changing the initial time

There would have been no loss of generality if in (6) we had taken the initial time t_0 to be 0, as the general case may be reduced to the case where $t_0 = 0$ by a change of variables $t \to t' + t_0$. Here we give formulas that express $\Gamma(\tau, \theta; \theta_0)$ in terms of $\Gamma(\tau, \theta - \theta_0; 0)$ and therefore allows one to express the coefficients $\alpha(t; t_0)$, $\bar{\beta}(t_0)$, ... in terms of the coefficients $\alpha(t; 0)$, $\bar{\beta}(0)$, ...

We introduce, for each $\theta \in \mathbb{T}^d$, the linear map $\Xi_\theta : \mathbb{C}^{\mathscr{W}} \to \mathbb{C}^{\mathscr{W}}$ defined as follows: given $\delta \in \mathbb{C}^{\mathscr{W}}$, $(\Xi_\theta \delta)_\emptyset = \delta_\emptyset$, and for each word $w = \mathbf{k}_1 \cdots \mathbf{k}_n$ with $n > 0$ letters,

$$(\Xi_\theta \delta)_{\mathbf{k}_1 \cdots \mathbf{k}_n} = e^{i(\mathbf{k}_1 + \cdots + \mathbf{k}_n) \cdot \theta} \delta_{\mathbf{k}_1 \cdots \mathbf{k}_n}.$$

Note that Ξ_θ is actually an algebra automorphism, as it preserves the convolution product: $\Xi_\theta(\delta \star \delta') = (\Xi_\theta \delta) \star (\Xi_\theta \delta')$, if $\delta, \delta' \in \mathbb{C}^{\mathscr{W}}$. In addition it maps $\mathscr{G}$ into $\mathscr{G}$.

The following result may be proved by induction on the number of letters using the recursive formulas (12), or, alternatively, by using the transport equation.

Proposition 3. *For each $\tau \in \mathbb{R}$ and $\theta, \theta_0 \in \mathbb{T}^d$,*

$$\Gamma(\tau, \theta; \theta_0) = \Xi_{\theta_0} \Gamma(\tau, \theta - \theta_0; 0).$$

As a consequence we have (cf. Theorem 3):

Corollary 1. *For arbitrary $\tau_1, \tau_2 \in \mathbb{R}$ and $\theta_1, \theta_2 \in \mathbb{T}^d$,*

$$\Gamma(\tau_1, \theta_1; 0) \star \Xi_{\theta_1} \Gamma(\tau_2, \theta_2; 0) = \Gamma(\tau_1 + \tau_2, \theta_1 + \theta_2; 0).$$

4 Autonomous problems

In this section consider a general class of perturbed autonomous problems. By building on the foundations laid down above we provide a method for reducing them to normal form.

4.1 Perturbed problems

We now study initial value problems

$$\frac{d}{dt}x = g(x) + f(x), \quad x(0) = x_0, \tag{21}$$

where $f, g : \mathbb{C}^D \to \mathbb{C}^D$. In the situations we have in mind, this system is seen as a perturbation of the system $(d/dt)x = g(x)$ whose solutions are known. In what follows we denote by g_j, $j = 1, \ldots, d$, a family of linearly independent vector fields that commute with each other (i.e. $[g_j, g_k] = 0$) and, for each $u = [u_1, \cdots, u_d] \in \mathbb{C}^d$, we set

$$g^u = \sum_{j=1}^{d} u_j g_j. \tag{22}$$

We always work under the following hypotheses:

- f may be decomposed as

$$f(x) = \sum_{\ell \in A} f_\ell(x) \tag{23}$$

  for a set of indices A, referred to as the alphabet as in the preceding sections.
- For each $j = 1, \ldots, d$ and each $\ell \in A$, there is $v_{j,\ell} \in \mathbb{C}$ such that

$$[g_j, f_\ell] = v_{j,\ell} f_\ell. \tag{24}$$

- There is $v \in \mathbb{C}^d$ such that $g = g^v$.
- The alphabet A is an additive monoid with neutral element $\mathbf{0}$,[2] such that, for each $j = 1, \ldots, d$ and $\ell, \ell' \in A$, $v_{j,\ell+\ell'} = v_{j,\ell} + v_{j,\ell'}$. In particular, $v_{j,\mathbf{0}} = 0$ for all j.
- The vector $v = (v_1, \ldots, v_d) \in \mathbb{C}^d$ is *non-resonant*, in the sense that, given $\ell \in A$, $v_1 v_{1,\ell} + \cdots + v_d v_{d,\ell} = 0$ if and only if $\ell = \mathbf{0}$.

The following proposition, whose proof may be seen in [13], shows that (24) may be reformulated in terms of the flows φ_u at time $t = 1$ of the vector fields g^u, $u \in \mathbb{C}^d$. Here and later, we use the notation

$$v_\ell^u = u_1 v_{1,\ell} + \cdots + u_d v_{d,\ell}.$$

Proposition 4. *Equation (24) is equivalent to the requirement that for each $x \in \mathbb{R}^D$, $u \in \mathbb{C}^d$, $\ell \in A$,*

$$\varphi_u'(x)^{-1} f_\ell(\varphi_u(x)) = \exp(v_\ell^u) f_\ell(x). \tag{25}$$

Before providing examples of systems that satisfy the hypotheses above, we shall obtain a word series representation of the solution of (21). Use the ansatz $x(t) = \varphi_{tv}(z(t))$ and invoke (25), to find that $z(t)$ must be the solution of

$$\frac{d}{dt} z = \sum_{\ell \in A} \exp(t v_\ell^v) f_\ell(z), \quad z(0) = x_0.$$

Since this problem is of the form (1) with

$$\lambda_\ell(t) = \exp(t v_\ell^v), \qquad \ell \in A, \tag{26}$$

---

[2] Recall that this means that A possesses a binary operation $+$ that is commutative and associative and such that $\mathbf{0} + \ell = \ell$ for each $\ell \in A$.

we find that $z(t) = W_{\alpha(t;0)}(x_0)$, where the coefficients $\alpha(t;0) \in \mathscr{G}$ are given by (3). In what follows, we will simply write $\alpha(t) = \alpha(t;0)$. Thus the solution of (21) has the representation

$$x(t) = \varphi_{tv}(W_{\alpha(t)}(x_0)). \tag{27}$$

Note that the coefficients $\alpha_w(t)$ *depend on v and the* $v_{j,\ell}$ *but are otherwise independent of g and* f_ℓ, $\ell \in A$.

Systems that satisfy the assumptions include the following (additional examples and further discussion may be seen in [13]):

Example 1. Consider systems of the form

$$\frac{d}{dt}x = Lx + f(x), \tag{28}$$

where L is a diagonalizable $D \times D$ matrix and each component of $f(x)$ is a power series in the components of x. Let $\mu_1, \ldots, \mu_d$ denote the distinct nonzero eigenvalues of L, so that L may be uniquely decomposed as

$$L = \mu_1 L_1 + \cdots + \mu_d L_d,$$

where the $D \times D$ matrices $L_1, \ldots, L_d$ are projectors ($L_j^2 = L_j$) with $L_j L_k = 0$ if $j \neq k$. Thus (22) holds for $g_j(x) = L_j x$, $v_j = \mu_j$. Furthermore (see [13] for details) f may be decomposed as $f = \sum_{\mathbf{k}} f_{\mathbf{k}}$, where the 'letters' $\mathbf{k}$ are elements of $\mathbb{Z}^d$, $\mathbf{k} = [k_1, \ldots, k_d]$, and, for each j and $\mathbf{k}$, $[L_j, f_{\mathbf{k}}] = k_j f_{\mathbf{k}}$. Analytic systems of differential equations having an equilibrium at the origin are of the form (28), provided that the linearization at the origin is diagonalizable; the perturbation f then contains terms of degree > 1 in the components of x.

As we shall point out later, Theorem 11 addresses the well-known problem, which goes back to Poincaré and Birkhoff [2], of reducing (28) to normal form.

Example 2. Consider next real systems of the form

$$\frac{d}{dt}\begin{bmatrix} y \\ \theta \end{bmatrix} = \begin{bmatrix} 0 \\ \omega \end{bmatrix} + f(y, \theta), \tag{29}$$

where $y \in \mathbb{R}^{D-d}$, $0 < d \leq D$, $\omega \in \mathbb{R}^d$ is a vector of frequencies $\omega_j \neq 0$, $j = 1, \ldots, d$, and θ comprises d angles, so that $f(y, \theta)$ is 2π-periodic in each component of θ with Fourier expansion

$$f(y, \theta) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \exp(i\mathbf{k} \cdot \theta) \, \hat{f}_{\mathbf{k}}(y).$$

After introducing the functions

$$f_{\mathbf{k}}(y, \theta) = \exp(i\mathbf{k} \cdot \theta) \, \hat{f}_{\mathbf{k}}(y), \qquad y \in \mathbb{R}^{D-d}, \, \theta \in \mathbb{R}^d,$$

the system takes the form (21) with $x = (y, \theta)$ and

$$g(y, \theta) = \begin{bmatrix} 0 \\ \omega \end{bmatrix}.$$

The decomposition (23) holds for the monoid $A = \mathbb{Z}^d$, and, if each $g_j(x)$ is taken to be a constant unit vector, then $g = g^v$, $v_j = \omega_j$ ($j = 1, \ldots, d$). In addition, (24) is satisfied with $v_{j,\mathbf{k}} = i k_j$, for each $j = 1, \ldots, d$ and each $\mathbf{k} = (k_1, \ldots, k_d) \in A$. Thus the nonresonance condition above (i.e., $v_1 v_{1,\ell} + \cdots + v_d v_{d,\ell} = 0$ if and only if $\ell = \mathbf{0}$) now becomes the well-known requirement that $k_1 \omega_1 + \cdots + k_d \omega_d = 0$ with integer k_j, $j = 1, \ldots, d$, only if all k_j vanish

In the particular case where the last d components of f vanish identically, the differential equations for θ yield $\theta = \omega t + \theta_0$, and (29) becomes a nonautonomous system for y of the form (6). Thus, the format (21) is a wide generalization of the format studied in the preceding section.

4.2 The transport equation. Normal forms

All the results obtained in Section 3.2 can be generalized to the case at hand. We shall omit the proofs of the results that follow when they may be obtained by adapting the corresponding proofs in Section 3.

We first provide recurrences to find the coefficients required in (27).

Theorem 7. *Given $\tau \in \mathbb{R}$, $u \in \mathbb{C}^d$, define, for each $w \in \mathcal{W}$, $\gamma_w(\tau, u) \in \mathbb{C}$ by means of the following recursions. $\gamma_\emptyset(\tau, u) = 1$, and for $r \geq 1$, $\ell_0 \in A \setminus \{\mathbf{0}\}$, and $\ell_1, \ldots, \ell_n \in A$,*

$$\gamma_{\ell_0}(\tau, u) = \frac{1}{v_{\ell_0}^v} \left(\exp(v_{\ell_0}^u) - 1 \right),$$

$$\gamma_{\mathbf{0}^r}(\tau, u) = \tau^r / r!,$$

$$\gamma_{\mathbf{0}^r \ell_0}(\tau, u) = \frac{\gamma_{\mathbf{0}^r}(\tau, u) \exp(v_{\ell_0}^u) - \gamma_{\mathbf{0}^{r-1} \ell_0}(\tau, u)}{v_{\ell_0}^v},$$

$$\gamma_{\ell_0 \ell_1 \cdots \ell_n}(\tau, u) = \frac{\gamma_{(\ell_0 + \ell_1) \ell_2 \cdots \ell_n}(\tau, u) - \gamma_{\ell_1 \cdots \ell_n}(\tau, u)}{v_{\ell_0}^v}, \tag{30}$$

$$\gamma_{\mathbf{0}^r \ell_0 \ell_1 \cdots \ell_n}(\tau, u) = \frac{\gamma_{\mathbf{0}^r (\ell_0 + \ell_1) \ell_2 \cdots \ell_n}(\tau, u) - \gamma_{\mathbf{0}^{r-1} \ell_0 \ell_1 \cdots \ell_n}(\tau, u)}{v_{\ell_0}^v}.$$

Then, for each $w \in \mathcal{W}$,

$$\alpha_w(t) = \gamma_w(t, tv).$$

The transport problem (cf. (13)) is:

$$\frac{\partial}{\partial \tau} \gamma(\tau, u) + v \cdot \nabla_u \gamma(\tau, u) = \gamma(\tau, u) \star B(u), \quad \gamma(0, 0) = \mathbb{1}. \tag{31}$$

where $B(u) \in \mathfrak{g}$ is defined as $B_\ell(u) = \exp(v_\ell^u)$, $\ell \in A$, and $B_w(u) = 0$ if the length of $w \in W$ is not 1.

Lemma 1 needs some adaptation to the present circumstances. We say that a complex-valued function is *polynomially smooth* if it is a linear combination of terms of the form $\tau^k \exp(v_\ell^u)$, $j = 1, 2, 3, \ldots$, $\ell \in A$. For each $w \in \mathscr{W}$, the function $\gamma_w : \mathbb{R} \times \mathbb{C}^d \to \mathbb{C}$ in Theorem 7 is clearly *polynomially smooth*.

Lemma 2. *Let the vector $v \in \mathbb{C}^d$ be nonresonant. If a polynomially smooth function $z : \mathbb{R} \times \mathbb{C}^d \to \mathbb{C}$ satisfies*

$$\frac{\partial}{\partial \tau} z(\tau, u) + v \cdot \nabla_u z(\tau, u) = 0, \quad z(0,0) = 0,$$

then $z(\tau, u)$ is identically zero.

Instead of Proposition 2 and Theorem 2, we now have the following result.

Theorem 8. *The function $\gamma(\tau, u)$ given in Theorem 7 is the unique solution of problem (31) such that each $\gamma_w : \mathbb{R} \times \mathbb{C}^d \to \mathbb{C}$, $w \in \mathscr{W}$, is polynomially smooth. Furthermore, for each $\tau \in \mathbb{R}$, $u \in \mathbb{C}^d$, the element $\gamma(\tau, u) \in \mathbb{C}^{\mathscr{W}}$ belongs to $\mathscr{G}$.*

Our next aim is to derive a result similar to Theorem 3. We need to introduce, for each $u \in \mathbb{C}^d$, the algebra map $\Xi_u : \mathbb{C}^{\mathscr{W}} \to \mathbb{C}^{\mathscr{W}}$ defined as follows: Given $\delta \in \mathbb{C}^{\mathscr{W}}$, $(\Xi_u \delta)_\emptyset = \delta_\emptyset$, and for each word $w = \ell_1 \cdots \ell_n$ with $n \geq 1$ letters,

$$(\Xi_u \delta)_{\ell_1 \cdots \ell_n} = \exp(v_{\ell_1 + \cdots + \ell_n}^u) \delta_{\ell_1 \cdots \ell_n}.$$

This generalizes the map Ξ_θ we used in Section 3.

Theorem 9. *For arbitrary $\tau_1, \tau_2 \in \mathbb{R}$ and $u_1, u_2 \in \mathbb{C}^d$,*

$$\gamma(\tau, u) \star (\Xi_u \gamma(\tau', u')) = \gamma(\tau + \tau', u + u').$$

Let us provide an interpretation of the last result in terms of maps in $\mathbb{C}^D$ (rather than in terms of elements of $\mathscr{G}$). In [13], it is proved that, for arbitrary $\delta \in \mathscr{G}$ and $u \in \mathbb{C}^d$

$$W_\delta(\varphi_u(x)) = \varphi_u(W_{\Xi_u \delta}(x)). \tag{32}$$

If we denote, for each $(\tau, u) \in \mathbb{R} \times \mathbb{C}^d$,

$$\Phi_{\tau, u}(x) = \varphi_u(W_{\gamma(\tau, u)}(x)),$$

then, for arbitrary $\tau, \tau' \in \mathbb{R}$ and $u, u' \in \mathbb{C}^d$,

$$\begin{aligned}
\Phi_{\tau, u}(\Phi_{\tau', u'}(x)) &= \varphi_u(W_{\gamma(\tau, u)}(\varphi_{u'}(W_{\gamma(\tau', u')}(x)))) \\
&= \varphi_u(\varphi_{u'}(W_{\Xi_u \gamma(\tau, u)}(W_{\gamma(\tau', u')}(x)))) \\
&= \varphi_{u+u'}(W_{\gamma(\tau + \tau', u + u')}(x)) \\
&= \Phi_{\tau + \tau', u + u'}(x).
\end{aligned}$$

We have successively used the definition of Φ, equation (32), Theorem 9, equation (4), and, again, the definition of Φ. To sum up, we have proved the following result, which generalizes Proposition 5.3 in [4].

Theorem 10. *For arbitrary* $\tau, \tau' \in \mathbb{R}$ *and* $u, u' \in \mathbb{C}^d$,

$$\Phi_{\tau,u} \circ \Phi_{\tau',u'} = \Phi_{\tau+\tau',u+u'}$$

Since, in view of (27), the solution $x(t)$ may be written as $x(t) = \Phi_{t,tv}(x_0)$, the theorem implies the representations

$$x(t) = \Phi_{0,tv}(\Phi_{t,0}(x_0)) = \Phi_{t,0}(\Phi_{0,tv}(x_0)). \tag{33}$$

The group property $\Phi_{t,0} \circ \Phi_{t',0} = \Phi_{t+t',0}$ implies that $\Phi_{t,0}(x) = W_{\gamma(t,0)}(x)$ is the t-flow of the autonomous system

$$\frac{d}{dt}X = W_{\bar{\beta}}(X),$$

where $\bar{\beta} \in \mathfrak{g}$ is given by

$$\bar{\beta} = \frac{d}{dt}\gamma(t,0)\bigg|_{t=0}. \tag{34}$$

Similarly, $\Phi_{0,tu} \circ \Phi_{0,t'u} = \Phi_{0,(t+t')u}$, implies that, for each fixed $u \in \mathbb{C}^d$, $\Phi_{0,tu}(x) = \varphi_{tu}(W_{\gamma(0,tu)}(x))$ is the t-flow of an autonomous system

$$\frac{d}{dt}x = \tilde{g}^u(x);$$

differentiation with respect to t of the flow at $t = 0$ reveals that

$$\tilde{g}^u(x) = g^u(x) + W_{\rho(u)}(x), \tag{35}$$

where $\rho(u) \in \mathfrak{g}$ is given by

$$\rho(u) = \frac{d}{dt}\gamma(0,tu)\bigg|_{t=0}.$$

Since, from Theorem 10, the flows $\Phi_{t,0}$, $\Phi_{0,tu}(x)$, $\Phi_{0,tu'}(x)$, $u, u' \in \mathbb{C}^d$, commute with one another, so do the corresponding vector fields $W_{\bar{\beta}}(X)$, $\tilde{g}^u(x)$, $\tilde{g}^{u'}(x)$. After invoking (33), we summarize our findings as follows.

Theorem 11. *The system in the initial value problem* (21) *may be rewritten in the form*

$$\frac{d}{dt}x = g(x) + f(x) = \tilde{g}^v(x) + W_{\bar{\beta}}(x),$$

where $\tilde{g}^v(x)$ and $\bar{\beta}$ are respectively given by (35) *and* (34). *The vector fields $\tilde{g}^v(x)$ and $W_{\bar{\beta}}(x)$ commute with each other, with $g(x) + f(x)$ and with $\tilde{g}^u(x)$ for arbitrary $u \in \mathbb{C}^d$.*

Note that the recursions defining $\gamma(t, u)$ in Theorem 7 give rise to similar recursions that allows us to conveniently compute the coefficients $\bar{\beta}, \rho(u) \in \mathfrak{g}$.

In the particular case where $A = \mathbb{Z}^d$ and the eigenvalues $v_{j,\ell}$ lie on the imaginary axis, Theorem 11 essentially coincides with Theorem 5.5 of [4]. The techniques in [4] are similar to those used here, but use B-series rather than word series. The general case of Theorem 11 was obtained by means of *extended words series* (see Section 5) in [13], where in addition it is shown that the vector fields $\tilde{g}^u(x)$ are conjugate to $g^u(x)$ by a map of the form $x \mapsto W_\delta(x)$, where $\delta \in \mathscr{G}$. The decomposition in Theorem 11 may be regarded as providing a *normal form*, where the original vector field is written as a vector field $\tilde{g}^v(x)$ that is conjugate to $g^v(x)$ perturbed by a vector field $W_{\bar{\beta}}(x)$ that commutes with $\tilde{g}^v(x)$.

Remark. For Hamiltonian problems, the commutation results in Theorem 11 allows us to write down explitly integrals of motion of the given problem. Details may be seen in [4] and [13].

5 Further extensions

In this section we study generalizations of the perturbed system in (21). Extended word series, introduced in [14], are a convenient auxiliary tool to study those generalizations.

5.1 Extended word series

Just as the study of systems of the form (1) leads to the introduction of word series via the representation (2), the expression (27) suggests the introduction of *extended word series*. Given the commuting vector fields g_j, $j = 1, \dots, d$ and the vector fields f_ℓ, $\ell \in A$ in the preceding section, to each $(v, \delta) \in \mathbb{C}^d \times \mathbb{C}^{\mathscr{W}}$ we associate its *extended word series [14], [13]:*

$$\overline{W}_{(v,\delta)}(x) = \varphi_v(W_\delta(x)).$$

With this terminology, the solution of (21) in (27) may be written as $x(t) = \overline{W}_{(tv,\alpha(t))}(x_0)$.

The symbol $\overline{\mathscr{G}}$ denotes the set $\mathbb{C}^d \times \mathscr{G}$. Thus, for each t, the solution coefficients $(tv, \alpha(t))$ provide an example of element of $\overline{\mathscr{G}}$. For $(u, \gamma) \in \overline{\mathscr{G}}$ and $(v, \delta) \in \mathbb{C}^d \times \mathbb{C}^{\mathscr{W}}$ we set

$$(u, \gamma) \bigstar (v, \delta) = (v + \delta_\emptyset u, \gamma \star \varXi_u \delta) \in \mathbb{C}^d \times \mathbb{C}^{\mathscr{W}}.$$

For this operation $\overline{\mathscr{G}}$ is a noncommutative group, with unit $\overline{\mathbb{1}} = (0, \mathbb{1})$; $(\mathbb{C}^d, \mathbb{1})$ and $(0, \mathscr{G})$ are subgroups of $\overline{\mathscr{G}}$. (In fact $\overline{\mathscr{G}}$ is an outer semidirect product of $\mathscr{G}$ and the additive group $\mathbb{C}^d$, as discussed in Section 3.2 of [14].)

By using (4) and (32), it is a simple exercise to check that the product $\bigstar$ has the following implication for the composition of the corresponding extended word series

$$\overline{W}_{(v,\delta)}\big(\overline{W}_{(u,\gamma)}(x)\big) = \overline{W}_{(u,\gamma)\bigstar(v,\delta)}(x), \qquad \gamma, \in \mathscr{G}, \ \delta \in \mathbb{C}^{\mathscr{W}}, \ u,v \in \mathbb{C}^d.$$

5.2 More general perturbed problems

We now generalize the problem (21), and allow a more general perturbation:

$$\frac{d}{dt}x = g(x) + W_\beta(x), \quad x(0) = x_0,$$

where $\beta \in \mathfrak{g}$. Clearly, the original problem (21) corresponds to the particular case where $\beta_\ell = 1$ for each $\ell \in A$, and $\beta_w = 0$ if the length of the word w is not 1. Other choices of β are of interest [14] when analyzing numerical integrators by means of the method of modified equations [17].

Proceding as in the derivation of (27), we find that the flow of (21) is given by

$$x(t) = \overline{W}_{(tv,\alpha(t))}(x(0)),$$

where $\alpha(t) \in \mathscr{G}$ is the solution of

$$\frac{d}{dt}\alpha(t) = \alpha(t) \star \Xi_{tv}\beta, \quad \alpha(0) = \mathbb{1}.$$

Moreover, $\alpha(t) = \gamma(t, tv)$, where $\gamma(\tau, u)$ is the unique polynomially smooth solution of the transport problem

$$\frac{\partial}{\partial \tau}\gamma(\tau, u) + v \cdot \nabla\gamma(\tau, u) = \gamma(\tau, u) \star \Xi_u\beta, \quad \gamma(0,0) = \mathbb{1},$$

which clearly generalizes (31). For each $(\tau, u) \in \mathbb{R} \times \mathbb{C}^d$, the element $\gamma(\tau, u)$ belongs to the group $\mathscr{G}$. Note that the recursions (30) are not valid for general β.

In analogy with Theorem 9, we have that, for arbitrary $\tau, \tau' \in \mathbb{R}$ and $u, u' \in C^d$,

$$(u, \gamma(\tau, u))\bigstar(\tau', \gamma(\tau', u')) = (u + u', \gamma(\tau + \tau', u + u')).$$

Theorems 10 and 11 hold true for general $\beta \in \mathfrak{g}$.

References

1. Alamo, A., Sanz-Serna, J.M.: A technique for studying strong and weak local errors of splitting stochastic integrators, submitted (arXiv 601.07335)
2. Arnold, V.I.: Geometrical Methods in the Theory of Ordinary Differential Equations, 2nd ed. Springer, New York (1988)
3. Chartier, P., Murua, A., Sanz-Serna, J.M.: Higher-order averaging, formal series and numerical integration I: B-series. Found. Comput. Math. **10**, 695–727 (2010)
4. Chartier, P., Murua, A., Sanz-Serna, J.M.: Higher-order averaging, formal series and numerical integration II: the quasi-periodic case. Found. Comput. Math., **12**, 471–508 (2012)
5. Chartier, P., Murua, A., Sanz-Serna, J.M.: A formal series approach to averaging: exponentially small error estimates. DCDS A **32**, 3009–3027 (2012)
6. Chartier, P., Murua, A., Sanz-Serna, J.M.: Higher-order averaging, formal series and numerical integration III: Error bounds. Found. Comput. Math. **15**, 591–612 (2015)
7. Fauvet, F., Menous, F.: Ecalle's arborification-coarborification transforms and Connes-Kreimer Hopf algebra, preprint (arXiv 1212.4740v2)
8. Jacobson, N.: Lie Algebras. Dover, New York (1979)
9. Kawski, M., Sussmann, H.J.: Nonommutative power series and formal Lie algebraic techniques in nonlinear control theory. In: Helmke, U., Pratzel-Wolters, D., and Zerz, E. (eds.) Operators, Systems, and Linear Algebra, pp. 111–118. Teubner, Stuttgart (1997)
10. Hairer, E., Wanner, G.: On the Butcher group and general multi-value methods, Computing **13**, 1–15 (1974)
11. Murua, A., The Hopf algebra of rooted trees, free Lie algebras and Lie series. Found. Comput. Math. **6**, 387–426 (2006)
12. Murua A., Sanz-Serna, J.M.: Vibrational resonance: a study with high-order word-series averaging, Applied Mathematics and Nonlinear Sciences **1**, 239-146 (2016)
13. Murua A., Sanz-Serna, J.M.: Computing normal forms and formal invariants of dynamical systems by means of word series. Nonlinear Analysis, **138**, 326–345 (2016)
14. Murua A., Sanz-Serna, J.M.: Word series for dynamical systems and their numerical integrators. Found. Comput. Math. to appear (arXiv:1502.05528)
15. Reutenauer, C.: Free Lie Algebras. Clarendon Press, Oxford (1993)
16. Sanders, J.A., Verhulst, F., Murdock, J.: Averaging Methods in Nonlinear Dynamical Systems (2nd. ed.). Springer, New York (2007)
17. Sanz-Serna, J.M., Calvo, M.P.: Numerical Hamiltonian Problems. Chapman and Hall, London (1994)
18. Sanz-Serna, J.M., Murua, A.: Formal series and numerical integrators: some history and some new techniques. In: Lei Guo, Zhi-Ming (eds.) Proceedings of the 8th International Congress on Industrial and Applied Mathematics (ICIAM 2015), pp. 311–331. Higher Education Press, Beijing, (2015)