



Méthodes de moyennisation stroboscopique appliquées aux équations aux dérivées partielles hautement oscillantes

Guillaume Leboucher

► To cite this version:

Guillaume Leboucher. Méthodes de moyennisation stroboscopique appliquées aux équations aux dérivées partielles hautement oscillantes. Equations aux dérivées partielles [math.AP]. Université de Rennes, 2015. Français. NNT : 2015REN1S121 . tel-01326270

HAL Id: tel-01326270

<https://theses.hal.science/tel-01326270>

Submitted on 3 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de
DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Mathématiques
Ecole doctorale Matisse

présentée par
Guillaume Leboucher

préparée à l'unité de recherche IRMAR (n°6625)
Institut de Recherche Mathématiques de Rennes

**Méthodes de
moyennisation
stroboscopique
appliquées aux
équations aux
dérivées partielles
hautement
oscillantes.**

**Thèse soutenue à Rennes
le 8 Décembre 2015**

devant le jury composé de :

François CASTELLA

Professeur à l'université de Rennes / *président*

Assyr ABDULLE

Professeur à l'école Polytechnique de Lausanne /
rapporteur

Hélène BARUCQ

Professeur à l'université de Pau / *rapporteur*

Gilles VILMART

Maître de conférence à l'université de Genève /
examineur

Mechthild THALHAMMER

Professeur à l'université d'Innsbruck / *examineur*

Florian MEHATS

Professeur à l'université de Rennes / *directeur de
thèse*

Philippe CHARTIER

Directeur de recherche à l'INRIA de Rennes /
co-directeur de thèse

UNIVERSITÉ DE RENNES 1

N° attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--

THÈSE

pour obtenir le grade de

DOCTEUR de l'Université de Rennes 1

Spécialité : **Mathématiques**

préparée au laboratoire **IRMAR**

dans le cadre de l'École Doctorale **MATISSE**

présentée et soutenue publiquement

par

Leboucher Guillaume

le 8 Décembre 2015

Titre:

Méthodes de moyennisation stroboscopique appliquées aux équations aux dérivées partielles hautement oscillantes.

Directeur de thèse: **Florian Méhats**

Co-directeur de thèse: **Philippe Chartier**

Jury

M. Assyr Abdulle,	Rapporteur
Mme. Hélène Barucq,	Rapporteur
Mme. Mechtild Thalhamer,	Examineur
M. Gilles Vilmart,	Examineur
M. François Castella,	Examineur
M. Florian Méhats,	Directeur de thèse
M. Philippe Chartier,	Co-directeur de thèse

Résumé

Cette thèse présente des travaux originaux dans le domaine des méthodes de moyennisation d'ordre élevé. On s'intéresse notamment à des procédures de moyennisation dite stroboscopique ou quasi-stroboscopique dans des espaces de Banach ou de Hilbert. Ces procédures sont ensuite appliquées à des exemples concrets : des équations d'évolutions hautement oscillantes.

Plus précisément, on montre dans un premier temps un résultat de moyennisation stroboscopique dans un espace de Banach où l'on obtient des estimations d'erreurs exponentielles. Ce théorème est ensuite appliqué sur deux équations des ondes semi-linéaire hautement oscillantes. On montre également que la *Stroboscopic Averaging Method* s'applique à une équation des ondes semi-linéaire avec conditions de Dirichlet. On trouve enfin numériquement, une dynamique intéressante de l'équation des ondes semi-linéaire mise en lumière par la procédure de moyennisation.

Dans un second temps, on présente un théorème de moyennisation quasi-stroboscopique dans un espace de Hilbert quelconque avec des estimations d'erreurs exponentielles. Ce théorème est alors appliqué de façon indirecte à une équation de Schrödinger semi-linéaire oscillante. Cette équation est d'abord projeté dans un espace de dimension finie pour qu'on puisse lui appliquer le théorème de moyennisation quasi-stroboscopique. On écrit alors un résultat de moyennisation quasi-stroboscopique pour l'équation de Schrödinger semi-linéaire avec des estimations d'erreur polynomiales.

Abstract

This thesis presents some original work in the field of high order averaging procedure. In particular, we are interested in stroboscopic and quasi-stroboscopic averaging procedure in abstract Banach or Hilbert spaces. This procedure is applied to concrete examples: some highly oscillatory evolution equations.

More precisely, we first show a theorem of stroboscopic averaging in a Banach space where we obtain exponential error estimates. This theorem is then applied on two semi-linear and highly oscillatory wave equations. We also put in evidence that the *Stroboscopic Averaging Method* works fine with a semi-linear wave equation with Dirichlet conditions. Finally, the averaging procedure puts in evidence, numerically, an interesting dynamics regarding the semi-linear wave equation with Dirichlet conditions.

In a second part, we present a quasi-stroboscopic averaging theorem in a Hilbert space with exponential error estimates. This theorem is applied on a semi-linear Schrödinger equation. This equation has first, to be projected in a finite dimensional space in order to fit in the hypotheses of the theorem. We then write a quasi-stroboscopic averaging theorem for a semi-linear Schrödinger equation with polynomial error estimates.

Remerciements

Je remercie toutes celles et ceux qui, de près ou de loin, ont contribué à la réalisation de cette thèse. Tout d'abord, ceux sans qui, la tenue d'une soutenance aurait été tout simplement impossible.

- Florian et Philippe, mes directeurs de thèse qui ont pris la peine de relire et corriger mes innombrables fautes (qu'elles soient d'Anglais ou Mathématiques) tout au long de ces quatre ans, qui m'ont aidé et guidé tant bien que mal et qui ne m'ont pas laissé tomber.
- Assyr Abdulle et Hélène Barucq, qui ont accepté de relire et commenter cette thèse.
- Mechtild Thalhammer, Gilles Vilmart et François Castella qui ont accepté de faire partie du jury et me permettent de leur exposer mes travaux.

Ensuite tout ceux qui ont été présent durant ces quatres ans afin de m'aider ou de m'inspirer:

- Chantal Halet, Carole Wosiak, Hélène Rousseaux et Marie-Aude Verger qui forment l'équipe de secrétaire la plus efficace avec lesquelles il m'ait été donné de travailler,
- L'équipe IPSOS de l'INRIA dont Philippe est le directeur,
- L'équipe d'analyse et d'analyse numérique de l'université de Rennes,
- L'équipe de sciences en cour[t]s qui m'a permis de tourner un court métrage parlant de mon sujet de thèse,
- Alexander Osterman, pour ses invitations à Innsbruck et à Dobbiaco,
- Anti, pour avoir rendu des conférences réellement passionantes.

En troisième lieu, je tiens à remercier les Rennais qui m'ont soutenu avant, pendant et après la thèse:

- Bastien, qui est le seul doctorant à organiser une raclette pour son pot de thèse,
- Florent, pour les délires et les soirées passées rue de la soif,
- Grand Ju, qui ne finira jamais de me faire rire,
- Cristophe, dont je pensais être débarrassé après ma thèse mais qui m'a suivi jusqu'à Paris,
- Yohann, excellent chercheur à qui il faudrait donner un poste, j'ai un pari en cours avec lui,
- Jérémy, aussi aimable et gentil qu'il est bon mathématicien,
- Quentin, qui me suit depuis la première et qui a poussé le vice jusqu'à soutenir le même jour que moi.
- Lisa, qui m'a toujours bien accueilli sur Rennes.
- Hélène, toujours présente, surtout quand je m'y attend le moins,
- Fif, auteur des paris les plus fous et qui pourtant n'a pas osé parier que je ne soutiendrais jamais (pour ce que j'en sais en tout cas),
- Manue et les jumeaux qui m'ont toujours accepté dans leurs vies (comme si ils avaient le choix) et m'ont permis de squatter chez eux quand je venais sur Rennes,
- Elise, qui m'a soutenu et motivé tout au long de la thèse (Et pas seulement pour aller courir).

Si il est important de connaître des matheux, des thésards afin de pouvoir échanger et discuter, je pense qu'il est tout aussi (voire plus) important d'avoir des amis en dehors de l'ENS et de l'IRMAR. J'aimerais les remercier ici même s'il semble peu probable qu'ils lisent un jour ces mots.

- Le personnel du collège Hector Berlioz, particulièrement disponibles et sympathiques. On ne pouvait pas imaginer pire comme environnement et rêver mieux comme collègues. Plus particulièrement, je remercie Teva, Laurent, Carole, Joan, Muriel, Juliette et Isabelle pour des raisons différentes.

- Les gens de Pokerennes qui m'ont accueilli dans leur petite famille. Plus particulièrement Bruno, Step, Vince, Nior et Koni avec qui mes échanges pokeristiques ont toujours été passionnant.
- Les manceaux qui sont avec moi depuis toujours : Samy, Lucile, Clément Thomas, Priscillia, Didi, Gaby, Sophie et la grande Emilie. Je leur dois beaucoup.
- Fabien, mon plus vieil ami qui même sans comprendre, osait me demander ce qu'il y avait dans ma thèse.

Enfin, je tiens à remercier tout ceux que je considère comme ma famille pour leur présence et l'aide infinie qu'ils m'ont apporté.

- Martin mon cousin Germain et Daniela.
- La famille Deichelborher avec Jean, Monique, Mireille, Pierrot, Loïc et Clément.
- Ma belle famille, notamment Jean-Marc, Chantal, Nicolas, Mathieu, Edwige, Eloïse et Aubin (mes premiers nièces et neveux) ainsi que le reste des Alsaciens.
- La famille Moreau avec Laurence, Christophe, Corentin, Laurie et Elise.
- Les Normands, Eric, Edith, Emeric, Martha, Emeline, Erika, Raphaël et Ethan.
- La famille Chaussavoine avec Annie, Jean-Luc, Alizée et Vincent.
- Mes colocs de toujours, Thomas et Juju.
- Ma grand mère, une femme qui m'impressionnera toujours.
- Laurent et Maud: Et oui je serais docteur avant vous finalement.
- Filou, Loris, Marie, Thibault, Amaury, Lancelot, Bénédicte, ainsi que mes parents. Je leur dois tellement à tous qu'ils seraient bien trop long de l'exposer ici.
- Mama.

Cette thèse est dédiée à mon grand père et à ma tante.

Contents

Résumé	iii
Abstract	iv
Remerciements	v
Contents	ix
1 Méthodes de moyennisation stroboscopique appliquées aux équations différentielles ordinaires	1
1 Introduction et historique	1
1.1 Un petit historique de la moyennisation	1
1.2 Problèmes de références	2
1.3 Exemples de problèmes pour les Equations Différentielles Ordinaires.	5
2 Différentes méthodes de résolutions.	9
2.1 Résolution utilisant les séries de Taylor.	10
2.2 Résolution utilisant les B-séries	12
2.3 Résolution utilisant une équation de transport	18
3 Une méthode numérique: La SAM (Stroboscopic Averaging Method)	22
3.1 Principe général de la SAM	23
3.2 Estimations d'erreurs et exemples numériques	24
4 Les problèmes quasi-périodiques	28
5 Résumé des résultats de la thèse	30
5.1 Différentes méthodes pour différents point de vue	30
5.2 Chapitre 2 : Méthode de moyennisation stroboscopique et équation des ondes hautement oscillantes	31
5.3 Chapitre 3 : Méthode de moyennisation quasi-stroboscopique dans des espaces de Hilbert	33
5.4 Chapitre 4 : Méthode de moyennisation quasi-stroboscopique pour une équation de Schrödinger semi-linéaire.	34
2 Stroboscopic averaging of highly-oscillatory non-linear wave equations	37
1 Introduction	37
2 High-order averaging and analytic functions in a Banach space	40
	ix

2.1	Preliminaries	40
2.2	Perko's formulas	41
2.3	Averaging theorem with polynomial remainder	43
2.4	Analytic functions in a Banach space	45
2.5	Averaging theorem with exponential error estimate	48
2.6	Geometric properties of stroboscopic averaging	50
3	Two examples : Wave equations with Dirichlet and periodic boundary conditions	53
3.1	A semi-linear wave equation with Dirichlet Boundary conditions.	53
3.2	A semi-linear wave equation with periodic boundary conditions.	58
4	Stroboscopic Averaging Method for a semi-linear wave equation.	61
4.1	The numerical method	61
4.2	A semi-linear wave equation with Dirichlet boundary conditions: Numerical results.	63
4.3	Energy exchange between modes in a non linear wave equation : heuristic analysis.	66
5	Appendix	69
3	Quasi-stroboscopic averaging in Hilbert spaces	71
1	Introduction	71
1.1	The formal equations of quasi-stroboscopic averaging	72
2	Abstract settings	74
2.1	Analytic functions in a Hilbert space	74
2.2	Quasi-periodic functions with values in $\mathbf{A}_\rho$	75
2.3	Analytic quasi-periodic functions with values in $\mathbf{A}_\rho$	76
2.4	Main Result: Exponential estimates for the quasi-stroboscopic averaging procedure	79
3	Technical proofs and intermediate results	80
3.1	Preliminary Lemmas	80
3.2	Resolution of (1.11) with polynomial remainder estimates	83
3.3	Proof of Theorem 2.13	85
4	Quasi-stroboscopic procedure for a semi-linear Schrödinger equation	89
1	Introduction	89
2	Projection of the nonlinear Schrödinger equation	91
2.1	Geometric form of the nonlinear Schrödinger equation	91
2.2	Hamiltonian character of equation (2.2)	94
2.3	Projection in a finite dimensional space	95
3	Application of the quasi-stroboscopic averaging procedure in X_L	97
3.1	A theorem of quasi-stroboscopic averaging	98
4	Quasi-periodic averaging with polynomial remainder estimates	103

Bibliography	107
--------------	-----

Chapter 1

Méthodes de moyennisation stroboscopique appliquées aux équations différentielles ordinaires

1 Introduction et historique

1.1 Un petit historique de la moyennisation

L' étude des problèmes perturbés vient originellement de l'objectif de relier les observations sur le déplacement des planètes avec les théories de Newton. Les premières tentatives des physiciens furent de créer des tables astronomiques pour chaque nouveau cas qu'ils rencontraient. Cependant aucune théorie globale ne permettait de résoudre des équations différentielles toujours plus complexe. L'idée de regarder un problème moyenné (souvent plus simple) que le problème de base afin d'en déduire des résultats qualitatifs vient notamment de Clairaut [21], Laplace [26] et Lagrange [36]. De nombreux autres scientifiques comme Jacobi, Van der Pol ou Poincaré, ont par la suite, repris cette méthode.

Cependant, la relation entre les solutions de l'équation de départ et de l'équation moyennée, bien que très utilisée en physique, restait formelle. La première preuve de validité asymptotique fut apporté par Fatou [28] en 1928 pour des champs de vecteurs périodiques. En union Soviétique, à la même époque des résultats similaires sont obtenus par Mandelstam et Papalexi [33]. Une percée importante est faite par Krylov et Bogoliubov [35] lorsqu'ils appliquent les résultats de la moyennisation aux champs de vecteurs presque périodiques. Les théories sur la moyennisation sont alors rassemblées dans un livre important écrit par Bogoliubov et Mitropolsky [5] en 1955 (traduit en Anglais en 1961 [4]) .

Ce sera le point de départ d'un grand nombre de travaux explorant plusieurs domaines des mathématiques. On peut citer notamment :

- Les formes normales de Birkoff qui consiste à mettre un système Hamiltonien sous une forme plus simple à intégrer.
- La théorie KAM du nom des trois mathématiciens qui en sont à l'origine: Kolmogorov [34], Arnold [2] et Moser [41].
- La moyennisation d'ordre élevé notamment présentée par Perko [44] en 1969 puis par Ellison, Saenz et Dumas [27] qui est au coeur des travaux de cette thèse.

De nombreux autres auteurs ont apportés des contributions dans ces domaines aux cours des dernières décennies et il serait fastidieux de tous les citer et de n'en oublier aucun. Cependant, un certain nombre sont revenues régulièrement dans mes recherches. A propos des formes normales de Birkhoff, on peut se référer à Bambusi [3]. Au sujet de la moyennisation, le livre de Sanders, Verhulst et Murdock [47] contient énormément d'informations. Enfin, Hairer Lubich et Wanner [32] reprennent la théorie KAM ainsi que de nombreux autres concepts (B-séries, Erreur de l'analyse rétrograde ...). Ils sont également à l'origine d'une méthode de moyennisation particulière appelée "modulated fourier expansion" [22].

Durant ces dernières années, Chartier, Méhats, Murua, Sanz Serna et Castella ([9, 10, 11, 13, 16, 17, 20]) ont développé l'idée de moyennisation stroboscopique et trouvé une méthode numérique efficace pour résoudre les problèmes avec champs de vecteurs périodiques. Nous présentons maintenant plus en détail le principe de la moyennisation stroboscopique.

1.2 Problèmes de références

On s'intéresse tout d'abord aux problèmes en temps long de la forme

$$\frac{dy^\varepsilon}{dt} = \varepsilon f(t, y^\varepsilon), \quad y^\varepsilon(0) = y_0 \quad (1.1)$$

On considère un entier $D \geq 1$ et on note $X = \mathbb{R}^D$ ou $\mathbb{C}^D$. Dans l'équation (1.1), $y^\varepsilon(t)$ est dans X , $\varepsilon > 0$ est un petit paramètre et f est une fonction régulière et P -périodique par rapport à la variable t . On note alors $\mathbb{T}$ le tore de période P , c'est à dire $\mathbb{T} = \mathbb{R}/(P\mathbb{Z})$. Ces problèmes sont appelés hautement oscillants et l'on cherche à les résoudre sur un intervalle de temps $[0, T/\varepsilon]$ où T est indépendant de ε .

Ces problèmes sont équivalents aux problèmes dits hautements oscillants que l'on peut formuler ainsi:

$$\frac{dy^\varepsilon}{d\tau} = f\left(\frac{\tau}{\varepsilon}, y^\varepsilon\right), \quad y^\varepsilon(0) = y_0. \quad (1.2)$$

avec f , une fonction régulière de $\mathbb{T} \times X$ à valeurs dans X . On cherche alors si (1.2) à une unique solution pour $\tau \in [0, T]$. Ces deux problèmes, mathématiquement semblables, s'appliquent à de nombreux problèmes concrets (cf section 1.3). Si dans la thèse, on étudiera indistinctement l'une ou l'autre forme, on présente dans cette introduction, la moyennisation stroboscopique pour des problèmes sous la forme (1.1) (problèmes en temps long).

Le principe de la moyennisation est de trouver un changement de variable $\Phi : \mathbb{T} \times X \rightarrow X$ régulier ainsi que le flot Ψ d'une équation autonome

$$\frac{d}{dt}\Psi(t, y_0) = \varepsilon G(\Psi(t, y_0); \varepsilon) \quad (1.3)$$

tel que la solution y^ε du problème (1.1) s'écrit

$$y^\varepsilon(t) = \Phi(t, (\Psi(t, y_0))), \quad \forall t \in [0, T/\varepsilon]. \quad (1.4)$$

La fonction G est alors appelée champs de vecteur moyenné de l'équation (1.1). Il dépend du champs de vecteur f et de ses différentes dérivées.

En pratique, l'expression (1.4) n'est obtenue que dans des cas très particuliers. (Par exemple, dans le cas où f est une application linéaire). En règle générale, on cherchera à montrer une version tronquée de la relation (1.4). C'est à dire, pour tout $n \in \mathbb{N}^*$, on trouve un changement de variable $\Phi^{[n]} : \mathbb{T} \times X \rightarrow X$ et le flot $\Psi^{[n]}$ d'une équation autonome

$$\frac{d}{dt}\Psi^{[n]}(t, y_0) = \varepsilon \left(\sum_{j=1}^n \varepsilon^{j-1} G_j(\Psi^{[n]}(t, y_0)) \right), \quad (1.5)$$

tel que

$$\left| y^\varepsilon(t) - \Phi^{[n]}(t, \Psi^{[n]}(t, y_0)) \right|_X = \mathcal{O}(\varepsilon^n), \quad \forall t \in [0, T/\varepsilon] \quad (1.6)$$

où $|\cdot|_X$ est la norme usuelle de $\mathbb{R}^D$ ou $\mathbb{C}^D$ et $\mathcal{O}$ est la notation de Landau classique. Dans la suite, on utilisera la notation classique

$$\forall t \in \mathbb{R} \quad \Phi(t) : y \in X \rightarrow X \ni \Phi(t, y).$$

Il a été prouvé que si, pour tout $n \in \mathbb{N}^*$, le champs de vecteurs

$$G^{[n]} := G_1 + \varepsilon G_2 + \dots + \varepsilon^{n-1} G_n$$

est unique, il y a en revanche une infinité de choix possible pour le changement de variable $\Phi^{[n]}$. Classiquement, on impose alors que le changement de variable soit de moyenne nulle, c'est à dire,

$$\langle \Phi \rangle = \frac{1}{P} \int_0^P \Phi(\theta) d\theta \equiv 0. \quad (1.7)$$

Une autre possibilité qui mène alors à la moyennisation dite stroboscopique, est d'imposer $\Phi(0) = \text{Id}_X$. Ce choix de moyennisation a plusieurs avantages. Premièrement, la moyennisation stroboscopique est la seule qui préserve la structure du problème de base (cf [16, 17] et partie 2.3). Ensuite, cette procédure mène à une méthode numérique: la Stroboscopic Averaging Method (SAM) (cf [10] et partie 3).

On définit l'ensemble

$$\mathbf{T}_S = \{Pk, k \in \mathbb{Z}\}. \quad (1.8)$$

Les éléments de $\mathbf{T}_S$ sont appelés les temps stroboscopiques et jouent un rôle important dans la procédure de moyennisation stroboscopique. En effet, la solution y^ε de l'équation (1.1) et la solution (formelle) Y de l'équation (1.3) coïncident sur ces temps stroboscopiques. Ecrit différemment,

$$\forall \mathbf{t} \in \mathbf{T}_S, \quad y(\mathbf{t}) = Y(\mathbf{t}).$$

Cette remarque est essentielle pour la construction de la SAM.

On notera que dans les cas où la fonction f de l'équation (1.1) est analytique par rapport à la variable y , on peut raffiner l'inégalité (1.6). En effet, il est possible de construire (cf [13] et Partie 2.3) un changement de variable Φ^ε et un champ de vecteur G^ε tel que le flot Ψ^ε de l'équation autonome

$$\frac{d}{dt} \Psi^\varepsilon(t, y_0) = \varepsilon G^\varepsilon(\Psi(t, y_0)),$$

satisfasse la relation

$$|y^\varepsilon(t) - \Phi^\varepsilon(t, \Psi^\varepsilon(t, y_0))|_X \leq C_1 e^{-\frac{C_2}{\varepsilon}}$$

où C_1 et C_2 sont deux constantes positives indépendantes de ε .

Enfin on s'intéresse également aux problèmes où la fonction f dans (1.1) n'est plus périodique en t mais quasi-périodique (cf [42, 38]). Cela englobe notamment tous les problèmes de la forme

$$\frac{dy^\varepsilon}{dt} = \varepsilon f(t\omega_1, \dots, t\omega_d, y^\varepsilon), \quad y^\varepsilon(0) = y_0. \quad (1.9)$$

Dans l'équation ci dessus, d est un entier plus grand que 2, la fonction $f : \mathbb{T}^d \times X \rightarrow X$ est supposée régulière et les réels $\omega_1, \dots, \omega_d$ sont non rationnellement dépendant. Plus précisément, ils devront vérifier une hypothèse de non-résonnance forte qui sera exposé dans la partie 4. On définit alors $\omega = (\omega_1, \dots, \omega_d)$, le vecteur des fréquences de l'équation (1.9). Par souci de simplifications, on notera pour $(t, y) \in \mathbb{R} \times X$,

$$f(t\omega, y) = f(t\omega_1, \dots, t\omega_d, y).$$

Par analogie avec la cas périodique, on cherche alors un changement de variable $\Phi^{[n]} : \mathbb{T}^d \times X \rightarrow X$ régulier ainsi que le flot $\Psi^{[n]}$ d'une équation autonome tel que

$$\left| y^\varepsilon(t) - \Phi^{[n]}(t\omega, \Psi^{[n]}(t, y_0)) \right| = \mathcal{O}(\varepsilon^n), \quad \forall t \in [0, T/\varepsilon]. \quad (1.10)$$

1.3 Exemples de problèmes pour les Equations Différentielles Ordinaires.

Dans cette partie, on donne trois exemples d'équations différentielles ordinaires entrant dans la catégorie des problèmes de références cités plus haut.

Un problème jouet

On considère le problème suivant,

$$\frac{dx}{dt} = -2i\pi\varepsilon^2 t e^{-2i\pi t} x^2, \quad x(0) = x_0 \in \mathbb{C}^D. \quad (1.11)$$

Cette équation est intéressante pour appréhender le concept de moyennisation stroboscopique. On vérifie en introduisant la variable $\tilde{x} = \varepsilon t$ et $y = \begin{pmatrix} \tilde{x} \\ x \end{pmatrix}$, que le problème (1.11) peut être mis sous la forme (1.1)

$$\frac{dy}{dt} = \begin{pmatrix} \varepsilon \\ -2i\pi\varepsilon \tilde{x} e^{-2i\pi t} x^2 \end{pmatrix} = \varepsilon f(t, y).$$

La fonction f est 1-périodique par rapport à t , régulière par rapport à y : L'ensemble des temps stroboscopiques est tout simplement $\mathbb{Z}$.

On connaît une solution de l'équation (1.11) donnée par

$$x(t) = \frac{1}{\frac{1}{x_0} - \varepsilon^2 t e^{-2i\pi t} + i\frac{\varepsilon^2}{2\pi} (e^{-2i\pi t} - 1)}.$$

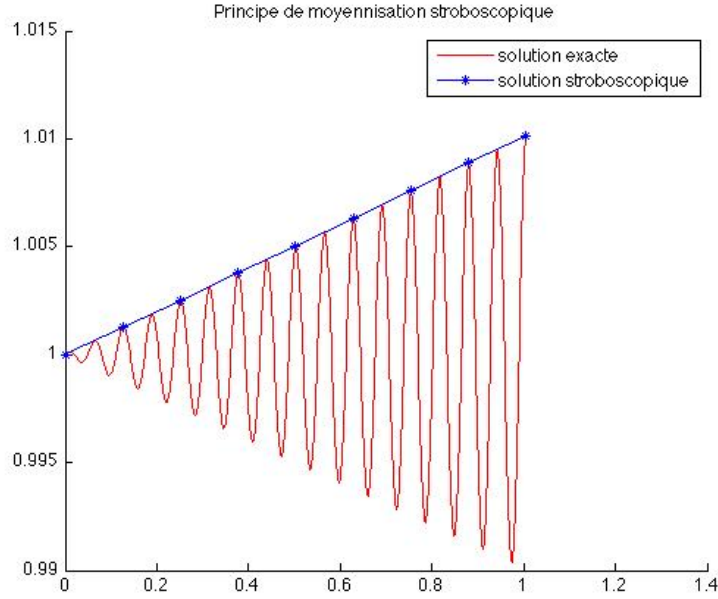


Figure 1.1: La solution de l'équation (1.11) et la solution de l'équation moyennée correspondante.

L'intérêt de cet exemple est qu'on peut calculer exactement le champ moyenné G^ε et donc la procédure de moyennisation stroboscopique est exacte. Après calculs, on trouve

$$G(z; \varepsilon) = \varepsilon z^2.$$

Ainsi le flot de l'équation

$$\frac{d\Psi^\varepsilon(t, x_0)}{dt} = \varepsilon G(\Psi^\varepsilon(t, x_0); \varepsilon) = \varepsilon^2 (\Psi^\varepsilon(t, x_0))^2$$

se calcule et l'on a

$$\Psi^\varepsilon(t, x_0) = \frac{1}{\frac{1}{x_0} - \varepsilon t}.$$

Il est plus fastidieux de calculer le changement de variable correspondant. Cependant, ce n'est pas nécessaire pour remarquer que pour $\mathbf{t} \in \mathbf{T}_s := \mathbb{Z}$, on a

$$x(\mathbf{t}) = \Psi^\varepsilon(\mathbf{t}, x_0).$$

On retrouve cette propriété sur la figure 1.1 où l'on voit que la courbe rouge (solution de l'équation initiale) et la courbe bleue (solution de l'équation moyennée) coïncident aux temps stroboscopiques.

L'équation de Van der Pol

L'équation de Van der Pol, du nom du physicien Néerlandais Balthasar Van der Pol est une équation hautement oscillante qui a trouvé de nombreuses applications en sciences physiques et biologiques. Étant donné $\omega_0 > 0$, l'oscillateur libre de Van der Pol est régi par l'équation

$$\ddot{x} + \omega_0^2 x = \varepsilon(1 - x^2)\dot{x}, \quad x(0) = x_0, \quad \dot{x}(0) = x_1. \quad (1.12)$$

Le changement de variable $y = \begin{pmatrix} x \\ \dot{x} \end{pmatrix}$ donne automatiquement l'équivalence entre l'équation (1.12) et l'équation suivante

$$\dot{y} = \begin{pmatrix} 0 & \omega_0 \\ -\omega_0 & 0 \end{pmatrix} y + \varepsilon F(y), \quad y(0) = \begin{pmatrix} x_0 \\ x_1 \end{pmatrix},$$

où la fonction F est donnée par

$$F : \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2 \rightarrow \mathbb{R}^2 \ni \begin{pmatrix} 0 \\ (1 - y_1^2)y_2 \end{pmatrix}.$$

Par souci de simplification, on définit $A := \begin{pmatrix} 0 & \omega_0 \\ -\omega_0 & 0 \end{pmatrix}$. Le deuxième changement de variable $v = e^{tA}y$ montre que l'équation (1.12) est équivalente à l'équation

$$\dot{v} = \varepsilon e^{tA} F(e^{-tA}v), \quad v(0) = \begin{pmatrix} x_0 \\ x_1 \end{pmatrix}$$

qui est bien sous la forme du problème (1.1).

Le problème de Fermi-Pasta-Ulam

Le problème de Fermi-Pasta-Ulam (ou problème FPU), du nom du Physicien Enrico Fermi, de l'informaticien John Pasta et du mathématicien Stanislas Ulam qui l'ont étudié [29] est un modèle simple de mécanique qui entraîne un comportement dynamique inattendu. On considère que m ressorts raides (partie linéaire) sont accrochés en alternance avec $m + 1$ ressorts mous (partie non linéaire). Les ressorts sont accrochés à chaque extrémités et on s'intéresse au mouvement de chaque points entre deux ressorts $x_1, \dots, x_{2m}$ (cf figure 1.3).

Pour tout $i = 1, \dots, 2m$, le point x_i satisfait l'équation différentielle suivante

$$\ddot{x}_i = (x_{i+1} + x_{i-1} - 2x_i) + \varepsilon((x_{i+1} - x_i)^2 - (x_i - x_{i-1})^2) \quad (1.13)$$

On note alors x le vecteur de $x_1, \dots, x_{2m}$. Le système d'équation associé au problème FPU s'écrit

$$\ddot{x} = Ax + \varepsilon g(x), \quad x(0) = x_0, \quad \dot{x}(0) = x_1, \quad (1.14)$$

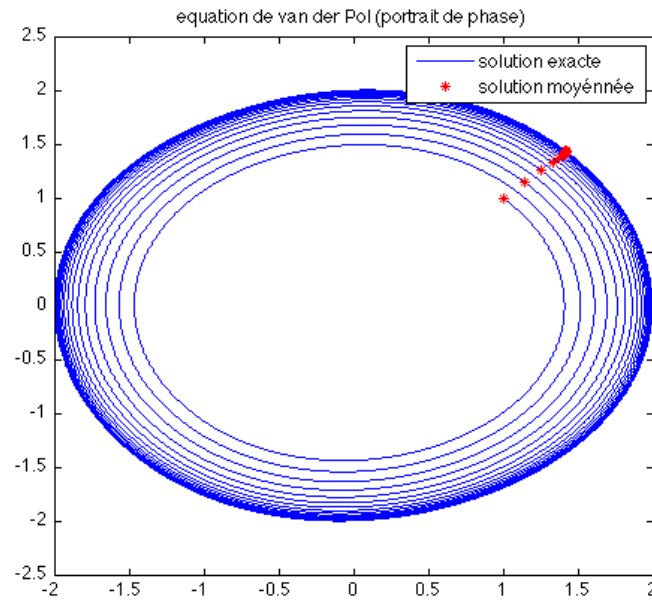


Figure 1.2: Portrait de phase de la solution de l'équation de Van der Pol (bleu) et de l'équation moyennée associée (rouge)

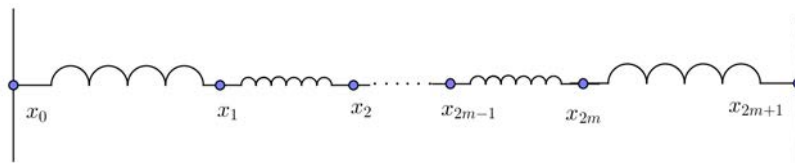


Figure 1.3: Le problème FPU.

où la matrice A est la matrice avec des -2 sur la diagonale et des 1 sur les sur et sous diagonales :

$$A := \begin{pmatrix} -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & \cdots & 1 & -2 & 1 & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \\ 0 & \cdots & 0 & 0 & 1 & -2 \end{pmatrix}$$

La matrice A est définie négative, on peut donc définir $\sqrt{-A}$. Appliquant le changement de variable $y = \begin{pmatrix} x \\ \dot{x} \end{pmatrix}$, on montre que le système (1.14) est équivalent au système

$$\dot{y} = \begin{pmatrix} 0 & \sqrt{-A} \\ -\sqrt{-A} & 0 \end{pmatrix} y + \varepsilon F(y), \quad y(0) = \begin{pmatrix} x_0 \\ x_1 \end{pmatrix}, \quad (1.15)$$

avec $F : y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in (\mathbb{R}^{2m})^2 \rightarrow \begin{pmatrix} 0 \\ g(y_1) \end{pmatrix}$.

On pose alors $B := \begin{pmatrix} 0 & \sqrt{-A} \\ -\sqrt{-A} & 0 \end{pmatrix}$ et le changement de variable $v = e^{tB} y$.

L'équation (1.14) est finalement équivalente à l'équation

$$\dot{v} = e^{tB} F(e^{-tB} v), \quad v(0) = \begin{pmatrix} x_0 \\ x_1 \end{pmatrix}. \quad (1.16)$$

On remarque alors que l'application $t \rightarrow e^{tB}$ est quasi-périodique avec un maximum de $2m$ fréquences (les quasi-périodes dépendent des valeurs propres de $\sqrt{-A}$). L'équation (1.16) est alors sous la même forme que l'équation (1.9) avec $d \leq 2m$.

2 Différentes méthodes de résolutions.

On reprend la propriété fondamentale de la moyennisation (1.4), En dérivant cette égalité par rapport au temps, on obtient

$$\frac{dy^\varepsilon}{dt}(t) = \partial_t \Phi(t, \Psi(t, y_0)) + \partial_y \Phi(t, \Psi(t, y_0))(\partial_t \Psi(t, y_0)).$$

On en déduit, d'après (1.1) et (1.3),

$$\varepsilon f(t, y^\varepsilon(t)) = \partial_t \Phi(t, \Psi(t, y_0)) + \varepsilon \partial_y \Phi(t, \Psi(t, y_0))(G^\varepsilon(\Psi(t, y_0))).$$

En posant alors $y = \Psi(t, y_0)$, on obtient une équation sur X ,

$$\partial_t \Phi(t, y) = \varepsilon f(t, \Phi(t, y)) - \varepsilon \partial_y \Phi(t, y) G^\varepsilon(y). \quad (2.1)$$

On sait qu'il n'existe pas en général de solution pour cette équation appelée équation homologique dans la théorie KAM. Néanmoins, plusieurs techniques permettent de trouver des quasi-solutions (le terme quasi-solution est introduit dans [13]) pour (2.1) que l'on présente maintenant.

2.1 Résolution utilisant les séries de Taylor.

Développement de Taylor des formules de la moyennisation

L'idée de base de la méthode est de présenter Φ et G comme des séries (formelles) en puissances de ε . Autrement dit, pour tout $t \in \mathbb{R}$, on suppose que

$$\Phi(t) = \text{Id} + \varepsilon \Phi_1(t) + \varepsilon^2 \Phi_2(t) + \dots \quad (2.2)$$

$$G = G_1 + \varepsilon G_2 + \varepsilon^2 G_3 + \dots \quad (2.3)$$

En admettant que le développement précédent soit valide, on a

$$\partial_t \Phi(t) = \varepsilon \partial_t \Phi_1(t) + \varepsilon^2 \partial_t \Phi_2(t) + \dots \text{ et } \partial_y \Phi(t) = \text{Id} + \varepsilon \partial_y \Phi_1(t) + \varepsilon^2 \partial_y \Phi_2(t) + \dots \quad (2.4)$$

On développe alors $f(t, \Phi(t, y))$ et $\partial_y \Phi(t, y) G^\varepsilon(y)$ en puissance de ε en utilisant leur séries de Taylor. Cela donne, pour les premiers termes,

$$f(t, \Phi(t, y)) = f(t, y) + \varepsilon \partial_y f(t, y) \Phi_1(t, y) + \mathcal{O}(\varepsilon^2) \quad (2.5)$$

et

$$\partial_y \Phi(t, y) G^\varepsilon(y) = G_1(y) + \varepsilon (G_2(y) + \partial_y \Phi_1(t, y) G_1(y)) + \mathcal{O}(\varepsilon^2). \quad (2.6)$$

On introduit alors les formules (2.4), (2.5) et (2.6) dans l'équation (2.1). En identifiant les puissances en ε , on obtient

$$\begin{aligned} f(t, y) &= \partial_t \Phi_1(t, y) + G_1(y). \\ \partial_y f(t, y) \Phi_1(t, y) &= \partial_t \Phi_2(t, y) + \partial_y \Phi_1(t, y) G_1(y) + G_2(y) \\ &\dots \end{aligned}$$

On remarque que pour $k \in \mathbb{N}^*$, ces équations sont toutes de la forme

$$\partial_t \Phi_k(t, y) + G_k(y) = g_k(t, y) \quad (2.7)$$

avec g_k une fonction ne dépendant que de f , $G_1, \dots, G_{k-1}$ et de $\Phi_1, \dots, \Phi_{k-1}$. De plus, g_k est P -périodique par rapport à t . La fonction Φ_k étant périodique en t , on prend la moyenne dans (2.7) et on obtient une formule pour G_k :

$$\forall k \in \mathbb{N}^*, \forall y \in X, \quad G_k(y) = \frac{1}{P} \int_0^P g_k(t, y) dt. \quad (2.8)$$

Le changement de variable Φ_k peut alors être calculé par la formule

$$\forall k \in \mathbb{N}^*, t \in \mathbb{R}, y \in X, \quad \Phi_k(t, y) = \int_0^t g_k(\theta, y) - G_k(y) d\theta. \quad (2.9)$$

Théorème de moyennisation de Perko

Pour tout n entier positif et toute région K convexe de X on définit $\mathcal{P}^n(K)$ l'ensemble des fonctions $f : [0, \infty[\times K \rightarrow X$ qui sont P -périodiques et continues par rapport à la première variable et qui sont de classe $\mathcal{C}^n$ par rapport à la deuxième variable.

On définit ensuite pour tout $t \in [0, \infty[$, l'approximation de y^ε d'ordre n

$$y_n(t) = Y_n(t) + \sum_{j=1}^n \varepsilon^j \Phi_j(t, Y_n(t)) \quad (2.10)$$

où la fonction Y_n est la solution du système autonome tronquée

$$\frac{dY_n}{dt} = G_1(Y_n) + \dots + \varepsilon^{n-1} G_n(Y_n). \quad (2.11)$$

Les fonctions $\Phi_1, \dots, \Phi_n, G_1, \dots, G_n$ sont alors respectivement données par les équations (2.9) et (2.8) avec les fonctions $g_1, \dots, g_n$ définis par (pour $j = 1, \dots, n$)

$$g_j(t, y) = \sum_{k=1}^{j-1} \left(\sum_{i_1 + \dots + i_k = j-1} \frac{1}{k!} \partial_y^k f(t, y) [\Phi_{i_1}, \dots, \Phi_{i_k}](t, y) - \partial_y \Phi_k(t, y) G_{j-k}(y) \right). \quad (2.12)$$

On redonne ici le théorème prouvé par Perko dans [44].

Theorem 2.1 Soient $T > 0$, $n \in \mathbb{N}^*$, K un domaine convexe de X , $y_0 \in K$, $f \in \mathcal{P}^n(K)$. On suppose que l'équation autonome d'ordre 1

$$\frac{dY_1}{d\tau} = \varepsilon G_1(Y_1), \quad Y_1(0) = y_0$$

a une unique solution $Y_1(t)$ incluse dans K pour tout $0 \leq t \leq T/\varepsilon$. Alors,

- Il existe $\varepsilon_n > 0$ tel que pour $0 < \varepsilon \leq \varepsilon_n$, l'équation (1.2)

$$\frac{dy^\varepsilon}{dt} = \varepsilon f(t, y^\varepsilon), \quad y^\varepsilon(0) = y_0$$

a une unique solution y^ε telle que $y(t) \in K$ pour $0 \leq t \leq T/\varepsilon$.

- L'approximation d'ordre n , notée y_n et définie par (2.10) satisfait pour $0 \leq t \leq T/\varepsilon$,

$$|y(t) - y_n(t)| \leq C_n \varepsilon^n \quad (2.13)$$

avec C_n une constante indépendante de ε .

Proof. On ne redonnera pas ici la preuve complète que l'on peut retrouver dans [44] mais les idées générales qui mènent au résultat.

La première étape est bien entendu de vérifier que pour $j = 1, \dots, n$ les fonctions g_j et Φ_j sont des fonctions de $\mathcal{P}^{n-j+1}(K)$ et donc que la formule des g_j est valide par récurrence.

La seconde étape est de montrer que puisque $Y_1(t)$ reste dans K pour tout $0 \leq \tau \leq T/\varepsilon$, l'équation autonome d'ordre n (2.11) a une unique solution Y_n qui reste également dans K lorsque ε est suffisamment petit.

La troisième et la plus importante étape est de calculer les restes dans les séries de Taylor de la différence $y^\varepsilon(t) - y_n(t)$ et de vérifier que ces restes sont bornés par ε^n .

Enfin, par un raisonnement en boucle (*bootstrap argument*), en supposant que l'estimation (2.13) est valide pour $t \leq T_1$, on montre qu'en choisissant ε suffisamment petit, la solution y^ε de l'équation (1.2) existe et reste dans K pour $0 \leq t \leq T/\varepsilon$.

■

Remark 2.2 Afin de faire le lien avec la partie 1.2, on écrit

$$\Phi^{[n]} = \text{Id} + \sum_{j=1}^n \varepsilon^j \Phi_j \quad \text{et} \quad G^{[n]} = \sum_{j=1}^n \varepsilon^{j-1} G_j.$$

2.2 Résolution utilisant les B-séries

En analyse numérique et notamment pour les méthodes de Runge-Kutta, il est intéressant de faire correspondre les différents termes de la solution d'une équation autonome développée en séries de Taylor avec des arbres à racine (*rooted tree*) [14, 40]. Cela permet de simplifier les notations, les calculs et donne des conditions d'ordre (souvent très simples) pour les méthodes numériques. Les B-séries sont alors des séries formelles sur l'ensemble des arbres à racines. La théorie des B-séries a été développée par Butcher dans les années 1963 - 1972 [7, 8] puis par Hairer et Wanner [31]. Le concept des B-séries a ensuite été appliquée par Chartier, Murua et Sanz-Serna [16] à la moyennisation stroboscopique.

Présentation des B-séries

Afin d'introduire les B-séries, on considère l'équation autonome suivante

$$\frac{dz}{dt} = \tilde{f}(z), \quad z(t_0) = z_0 \tag{2.14}$$

Le but ici est de trouver une méthode simple et efficace pour calculer les dérivées q-ème $\frac{d^q z}{dt^q}$ de la solution de cette équation. On a

$$\begin{aligned}\frac{dz}{dt} &= \tilde{f}(z) \\ \frac{d^2 z}{dt^2} &= \tilde{f}'(z) \frac{dz}{dt} \\ \frac{d^3 z}{dt^3} &= \tilde{f}''(z) \left(\frac{dz}{dt}, \frac{dz}{dt} \right) + \tilde{f}'(y) \frac{d^2 z}{dt^2} \\ \frac{d^4 z}{dt^4} &= \tilde{f}^{(3)} \left(\frac{dz}{dt}, \frac{dz}{dt}, \frac{dz}{dt} \right) + 3\tilde{f}''(z) \left(\frac{d^2 z}{dt^2}, \frac{dz}{dt} \right) + \tilde{f}'(y) \frac{d^3 z}{dt^3} \\ &\dots\end{aligned}$$

On remplace alors les dérivées de z dans les membres de droites afin d'obtenir des relations qui ne dépendent que des dérivées de $\tilde{f}$. On trouve formellement (pour des soucis de clarté, on a omis les dépendances en z),

$$\begin{aligned}\frac{dz}{dt} &= \tilde{f} \\ \frac{d^2 z}{dt^2} &= \tilde{f}' \tilde{f} \\ \frac{d^3 z}{dt^3} &= \tilde{f}'' (\tilde{f}, \tilde{f}) + \tilde{f}' \tilde{f}' \tilde{f} \\ \frac{d^4 z}{dt^4} &= \tilde{f}^{(3)} (\tilde{f}, \tilde{f}, \tilde{f}) + 3\tilde{f}'' (\tilde{f}' \tilde{f}, \tilde{f}) + \tilde{f}' \tilde{f}'' (\tilde{f}, \tilde{f}) + \tilde{f}' \tilde{f}' \tilde{f}' \tilde{f} \\ &\dots\end{aligned}$$

L'idée générale des B-séries est d'associer à chaque éléments différentiels qui se trouvent dans les membres de droites ci dessus, un arbre à racine.

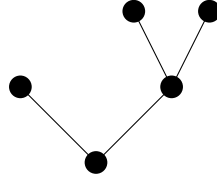
Definition 2.1 *L'ensemble des arbres à racines, noté $\mathcal{T}$, est défini récursivement par*

1. *Le graphe $\bullet$ muni d'un unique point (appelée la racine) appartient à $\mathcal{T}$.*
2. *Si $u_1, \dots, u_m \in \mathcal{T}$, alors le graphe obtenu en reliant les racines de $u_1, \dots, u_m$ à un point appartient encore à $\mathcal{T}$ noté $u = [u_1, \dots, u_m]$. Le nouveau point d'attache est la racine de u .*

On appelle alors l'ordre de l'arbre u et on note $|u|$ le nombre de points du graphe. Dans le tableau 2.2, on répertorie tous les arbres à racines d'ordre inférieur à 4.

On définit ensuite les éléments différentiels qui à chaque arbre lui associe les dérivées d'une fonction.

Figure 1.4: Exemple d'arbres à racines



Definition 2.2 Soit $f : X \rightarrow X$ une fonction régulière. L'élément différentiel d'un arbre à racines $u \in \mathcal{T}$ est une fonction $F(u) : X \rightarrow X$ défini par récurrence:

1. $F(\bullet)(y) = \tilde{f}(y)$.
2. $F(u)(y) = \frac{d^m}{dy^m} \tilde{f}(y)[F(u_1)(y), \dots, F(u_m)(y)]$ si on a $u = [u_1, \dots, u_m]$.

$ u $	u	Graphe	$F(u)$	$\sigma(u)$
1	$\bullet$		$\tilde{f}$	1
2	$[\bullet]$		$\tilde{f}' \tilde{f}$	1
3	$[[\bullet]]$		$\tilde{f}' \tilde{f}' \tilde{f}$	1
	$[\bullet, \bullet]$		$\tilde{f}''(\tilde{f}, \tilde{f})$	2
4	$[[[\bullet]]]$		$\tilde{f}' \tilde{f}' \tilde{f}' \tilde{f}$	1
	$[[\bullet, \bullet]]$		$\tilde{f}' \tilde{f}''(\tilde{f}, \tilde{f})$	2
	$[[\bullet], \bullet]$		$\tilde{f}''(\tilde{f}' \tilde{f}, \tilde{f})$	1
	$[\bullet, \bullet, \bullet]$		$\tilde{f}^{(3)}(\tilde{f}, \tilde{f}, \tilde{f})$	6

Table 1.1: Arbres à racines d'ordre inférieur à 4.

Si z est la solution exacte de l'équation (2.14), sa dérivé n -ième est donnée par la proposition suivante.

Proposition 2.3 Si la fonction $\tilde{f}$ est de classe $\mathcal{C}^m(X)$, la solution de l'équation (2.14) vérifie pour $q \leq m$,

$$\frac{d^q z}{dt^q}(t_0) = \sum_{|u|=q} \alpha(u) F(u)(z_0), \quad (2.15)$$

où les coefficients $\alpha(u)$ sont des entiers positifs.

Afin de trouver l'ordre d'une méthode numérique, on développe la solution exacte à l'aide de série de Taylor et l'on compare la solution numérique développée elle aussi en série. Ainsi, on a pour h suffisamment petit,

$$\begin{aligned} z(t_0 + h) &= z(t_0) + h \frac{dz}{dt}(t_0) + \frac{h^2}{2!} \frac{d^2z}{dt^2}(t_0) + \dots \\ &= y_0 + hf(y_0) + \frac{h^2}{2!} f'(y_0)f(y_0) + \dots \end{aligned}$$

L'introduction des B-séries est motivée par la formule précédente. Il s'agit de séries formelles en puissance de h utilisant les éléments différentiels $F(u)$ définis plus haut. Cependant, pour des raisons pratiques, on introduit également un coefficient de symétrie lié à chaque arbre.

Definition 2.3 Le coefficient de symétrie $\sigma(u)$ est défini par récurrence. On a $\sigma(\bullet) = 1$ et pour $u = [u_1, \dots, u_m]$,

$$\sigma(u) = \sigma(u_1) \dots \sigma(u_m) \mu_1! \mu_2! \dots,$$

où les entiers $\mu_1, \mu_2, \dots$ compte le nombre d'arbres identiques parmi $u_1, \dots, u_m$.

Par exemple, le coefficient de symétrie de l'arbre $[\bullet, \bullet]$ est 2. On explique plus en détail dans [32] la raison de l'introduction de ces coefficients de symétrie. Cela nous permet de donner la forme courante des B-séries.

Definition 2.4 Pour une application $a : \mathcal{T} \cap \{\emptyset\} \rightarrow \mathbb{R}$, on appelle B-séries une série formelle de la forme

$$B(a, z) = a(\emptyset)z + \sum_{u \in \mathcal{T}} \frac{h^{|u|}}{\sigma(u)} a(u) F(u)(z). \quad (2.16)$$

B-séries et formules de moyennisation

Dans cette partie, on va utiliser les B-séries dans la procédure de moyennisation stroboscopique. Revenons au problème en temps long (1.1)

$$\frac{dy^\varepsilon}{dt} = \varepsilon f(t, y^\varepsilon), \quad y^\varepsilon(0) = y_0$$

La fonction f étant P -périodique, on peut la développer en séries de Fourier. Ainsi on obtient

$$\frac{dy^\varepsilon}{dt} = \sum_{k \in \mathbb{Z}} e^{\frac{2i\pi}{P} kt} f_k(y^\varepsilon).$$

Afin de ne pas compliquer inutilement les calculs, on va supposer que dans cette partie, la fonction f est 2π -périodique. Ainsi le développement en série de Fourier donne,

$$\frac{dy^\varepsilon}{dt} = \sum_{k \in \mathbb{Z}} e^{ikt} f_k(y^\varepsilon).$$

On introduit alors ce qu'on appelle des arbres à racines colorés, c'est à dire des arbres où chaque point possède une étiquette $k \in \mathbb{Z}$.

Definition 2.5 *L'ensemble $\mathcal{U}$ des arbres à racines colorés se définit par récurrence.*

- Pour tout $k \in \mathbb{Z}$ le point coloré $(\bar{k})$ appartient à $\mathcal{U}$.
- Si $u_1, \dots, u_m \in \mathcal{U}$ alors le fait de greffer les racines de $u_1, \dots, u_m$ à un nouveau point de couleur k se note

$$u = [u_1, \dots, u_m]_k$$

et est un arbre de $\mathcal{U}$.

On note, comme pour les arbres à racines, $|u|$ le nombre de points du graphe. Les éléments différentiels sont définis en fonction des modes de Fourier de la fonction f .

Definition 2.6 *L'élément différentiel d'un arbre à racines coloré $u \in \mathcal{U}$ est une fonction $F(u) : X \rightarrow X$ défini par récurrence:*

1. $F(\bar{k})(y) = f_k(y)$.
2. Si on a $u = [u_1, \dots, u_m]_k$, on définit

$$F(u)(y) = f_k^{(m)}(y)[F(u_1)(y), \dots, F(u_m)(y)].$$

Les coefficients de symétrie $\sigma(u)$ sont définis de la même façon que précédemment (voir Définition 2.3). On a alors le résultat suivant: La solution y^ε de l'équation (1.2) possède un développement en B-séries

$$y^\varepsilon(t) = y_0 + \sum_{u \in \mathcal{U}} \varepsilon^{|u|} \frac{\beta_u(t)}{\sigma(u)} F(u)(y_0) = y_0 + \sum_{j \geq 1} \varepsilon^j \sum_{|u|=j} \frac{\beta_u(t)}{\sigma(u)} F(u)(y_0). \quad (2.17)$$

Les entiers $\beta_u(t)$ sont les coefficients élémentaires de l'arbre u et sont définis ci-après. (Les détails de l'obtention de cette série ainsi que les calculs des coefficients de β_u peuvent être trouvés dans [16]).

Definition 2.7 *Les coefficients $\beta_u(t)$ sont définis par récursivité.*

1. Pour l'arbre $u = (\underline{k})$, on a

$$\beta_u(t) = \int_0^t e^{ik\theta} d\theta.$$

2. Si u est de la forme $u = [u_1, \dots, u_m]_k$, on a

$$\beta_u(t) = \int_0^t e^{ik\theta} \beta_{u_1}(\theta) \dots \beta_{u_m}(\theta) d\theta.$$

Ce qui est intéressant ici, c'est que les coefficients $\beta_u(t)$ sont universels dans le sens qu'ils ne dépendent pas de la fonction f de l'équation de départ. Il ne semble pas possible d'obtenir des formules simple pour ces coefficients. Cependant, Chartier & al. montrent qu'il y a un moyen d'obtenir ces coefficients par récurrence. On introduit pour tout $r \in \mathbb{N}$ et $k \in \mathbb{Z}$, les fonctions

$$\phi_{r,k}(t) = t^r \exp(ikt) \quad \text{et} \quad \tilde{\beta}_{r,k}(t) = \int_0^t \phi_{r,k}(\theta) d\theta.$$

Les coefficients $\tilde{\beta}_{r,k}$ se calculent eux par la formule de récurrence suivante valide pour $k \neq 0$ et $r \in \mathbb{N}^*$,

$$\tilde{\beta}_{0,k}(t) = \frac{i}{k} - \frac{i}{k} e^{ikt}, \quad \tilde{\beta}_{r,k}(t) = \frac{ir}{k} \tilde{\beta}_{r-1,k}(t) - \frac{i}{k} \phi_{r,k}(t).$$

Quand $k = 0$, on a

$$\tilde{\beta}_{r,0}(t) = \frac{t^{r+1}}{r+1}.$$

On sait alors qu'il existe des constantes $a_{u,r,k} \in \mathbb{C}$ tel que les coefficients $\beta_u(\tau)$ s'écrivent

$$\beta_u(\tau) = \sum_{r \in \mathbb{N}, k \in \mathbb{Z}} a_{u,r,k} \phi_{r,k}(\tau) = \sum_{r \in \mathbb{N}, k \in \mathbb{Z}} a_{u,r,k} \tau^r e^{ik\tau}.$$

Les constantes $a_{u,r,k}$ peuvent alors être trouvées par récurrence en utilisant les fonctions $\tilde{\beta}_{r,k}$ (voir [16] pour plus d'informations).

En utilisant la décomposition en Fourier et les B-séries, on peut directement écrire la solution Y de l'équation moyennée

$$\frac{dY}{dt} = \varepsilon G(Y; \varepsilon)$$

sous la forme d'une B-série. Si l'on définit

$$\tilde{\alpha}_u(t) = \sum_{r \in \mathbb{N}, k \in \mathbb{Z}} a_{u,r,k} t^r,$$

alors Y s'écrit formellement comme

$$Y(t) = y_0 + \sum_{j=1} \varepsilon^j \sum_{|u|=j} \frac{\tilde{\alpha}_u(t)}{\sigma(u)} F(u)(y_0).$$

De là, on peut calculer les valeurs du champ moyen G en dérivant l'expression précédente par rapport à t .

Il faut préciser que les formules obtenues ici sont formelles dans le sens que les séries considérées ne sont pas convergentes. Afin d'obtenir des résultats concrets, on tronque les séries et on borne les erreurs associées. Les calculs et les bornes d'erreurs ont été obtenus par Chartier & al. [18] dans un cadre plus général (un cadre quasi-périodique, cf partie 4).

2.3 Résolution utilisant une équation de transport

On reprend l'équation (2.1) obtenue en dérivant l'égalité de moyennisation (1.4):

$$\varepsilon f(t, \Phi(t, y)) = \partial_t \Phi(t, y) + \varepsilon \partial_y \Phi(t, y) G(y).$$

On rappelle la notation de la moyenne

$$\langle \Phi(\cdot, y) \rangle = \frac{1}{P} \int_0^P \Phi(\theta, y) d\theta,$$

et on note $f \circ \Phi(t, y) = f(t, \Phi(t, y))$. On remarque alors qu'en prenant la moyenne dans l'équation (2.1) et en supposant que $\partial_y \Phi(t, y)$ est inversible, on obtient pour le champ de vecteur G^ε ,

$$G(y) = \langle \partial_y \Phi(\cdot, y) \rangle^{-1} \langle f \circ \Phi(\cdot, y) \rangle. \quad (2.18)$$

On en déduit alors l'équation de transport pour Φ

$$\partial_t \Phi(t, y) = \varepsilon (f(t, \Phi(t, y)) - \partial_y \Phi(t, y) \langle \partial_y \Phi(\cdot, y) \rangle^{-1} \langle f \circ \Phi(\cdot, y) \rangle). \quad (2.19)$$

On introduit ensuite deux opérateurs non-linéaires Λ et Γ^ε définis par

$$\Lambda(\Phi)(t, y) = f(t, \Phi(t, y)) - \partial_y \Phi(t, y) \langle \partial_y \Phi(\cdot, y) \rangle^{-1} \langle f \circ \Phi(\cdot, y) \rangle \quad (2.20)$$

et

$$\Gamma^\varepsilon(\Phi)(t, y) = y_0 + \varepsilon \int_0^t \Lambda(\Phi)(\theta, y) d\theta. \quad (2.21)$$

Avec ces notations l'équation (2.19) se réécrit

$$\partial_t \Phi(t, y) = \varepsilon \Lambda(\Phi)(t, y), \quad \Phi(0, y) = y \quad (2.22)$$

ou bien dans sa version intégrale

$$\Phi(t, y) = \Gamma^\varepsilon(\Phi)(t, y). \quad (2.23)$$

Au vu des résultats précédents, il n'est pas possible, en général de trouver une solution pour l'équation (2.22). On cherche alors des quasi-solutions qui ont par exemple été introduites dans [13].

Construction des quasi-solutions et estimation exponentielle

La construction de la moyennisation stroboscopique par l'équation de transport (2.22) n'a jamais été publié dans le cadre des EDOs. Actuellement, la seule publication sur le sujet est [13] qui donne la construction dans le cadre plus général des espaces de Banach (en dimension infinie) et donne l'exemple d'une équation de Schrödinger semi-linéaire. C'est cette construction que l'on présente succinctement ici.

Dans cet article, Castella & al. montrent également que lorsque la fonction f de l'équation (1.1) est analytique par rapport à la seconde variable (et non plus de classe $\mathcal{C}^n$ comme dans le Théorème 2.1) l'erreur dans l'estimation (2.13) est exponentiellement petite (par opposition à l'estimation en $\mathcal{O}(\varepsilon^n)$).

Ce type d'estimation est classique en théorie KAM (on pourra consulter à ce sujet [32]). En ce qui concerne la moyennisation stroboscopique, de premiers résultats ont été obtenus par Chartier & al. en utilisant les B-séries (cf [16]).

Pour tout ouvert K de $\mathbb{R}^D$, on définit l'élargissement de K dans $\mathbb{C}^D$ de taille $\rho > 0$ par

$$K_\rho = \{y + \tilde{y}, (y, \tilde{y}) \in K \times \mathbb{C}^D, \text{ s.t. } |\tilde{y}|_{\mathbb{C}^D} \leq \rho\}$$

On s'intéresse ensuite à l'ensemble $\mathbf{A}_\rho$ des fonctions $\varphi : (t, y) \in \mathbb{T} \times K_\rho \rightarrow \mathbb{C}^D \ni \tilde{\varphi}(t, y)$ analytiques et bornées sur K_ρ . L'ensemble A_ρ est alors un espace de Banach muni de la norme

$$\|\varphi\|_\rho = \sup_{(t, y) \in \mathbb{T} \times K_\rho} \|\varphi(t, y)\|_\rho.$$

On considère $n \in \mathbb{N}^*$ et la suite de fonctions

$$\Phi^{[0]}(t) = \text{Id}_X, \quad \Phi^{[k+1]}(t) = \Gamma^\varepsilon(\Phi^{[k]})(t), \quad k = 0, \dots, n, \quad (2.24)$$

(l'opérateur Γ^ε étant défini par (2.21)) ainsi que le champ de vecteur

$$G^{[k]}(y) = \langle \partial_y \Phi^{[k]}(\cdot, y) \rangle^{-1} \langle f \circ \Phi^{[k]}(\cdot, y) \rangle. \quad (2.25)$$

Pour finir, les fonctions G_k ($k = 0, \dots, n$) intervenants dans l'équation autonome (2.11) sont données par

$$G_k(y) = \frac{1}{k!} \frac{dG^{[k]}}{d\varepsilon^k} \Big|_{\varepsilon=0} (y) \quad (2.26)$$

On a alors le théorème suivant:

Theorem 2.4 *Soient $T > 0, R > 0$, K un domaine convexe de $\mathbb{R}^D$ et $y_0 \in K$. On suppose que la fonction $f : \mathbb{T} \times \mathbb{R}^D \rightarrow \mathbb{R}^D$ de l'équation (1.1) est bornée et analytique dans K_{2R} .*

- Alors, pour tout $n \in \mathbb{N}$, il existe $\varepsilon(n) > 0$, tel que pour $0 < \varepsilon < \varepsilon(n)$ et $k = 0, \dots, n$ les fonctions $\Phi^{[k]}$ et $G^{[k]}$ sont bien définies, P -périodique et de classe $\mathcal{C}^1$ par rapport à t et analytiques en la variable y .

De plus, il existe ε^* et pour $0 < \varepsilon \leq \varepsilon^*$, $n_\varepsilon \in \mathbb{N}$ tel que les propositions suivantes soient vérifiées (Par souci de simplification, on note $\Phi^\varepsilon = \Phi^{[n_\varepsilon]}$, $G^\varepsilon = G^{[n_\varepsilon]}$).

1. L'équation (1.1)

$$\frac{dy^\varepsilon}{dt} = \varepsilon f(t, y^\varepsilon), \quad y^\varepsilon(0) = y_0$$

a une unique solution y^ε satisfaisant $y^\varepsilon(t) \in K$ pour $0 \leq t \leq T/\varepsilon$.

2. La solution y^ε satisfait

$$\forall t \in [0, T/\varepsilon], \quad y^\varepsilon(t) = \Phi^\varepsilon(t, Y(t))$$

où Y est la solution de l'équation quasi-autonome suivante

$$\frac{dY}{dt} = \varepsilon G^\varepsilon(Y) + \mathcal{R}(t, Y). \quad (2.27)$$

Le reste $\mathcal{R}$ ci-dessus est alors $\mathcal{C}^1$ et périodique en t , analytique sur K_R par rapport à la seconde variable et vérifie

$$\|\mathcal{R}\|_R \leq C_1 e^{-C_2/\varepsilon}$$

où C_1 C_2 sont deux constantes positives indépendantes de ε .

3. On introduit alors Ψ^ε le flot de l'équation autonome

$$\frac{dY}{dt} = \varepsilon G^\varepsilon(Y), \quad Y(0) = y_0.$$

La solution y^ε de l'équation (1.1) est alors exponentiellement proche de $\Phi^\varepsilon \circ \Psi^\varepsilon(t, y_0)$. C'est à dire qu'il y a $\tilde{C}_1$ et $\tilde{C}_2$ deux constantes positives telles que

$$\forall t \in [0, T/\varepsilon], \quad |y(t) - \Phi^\varepsilon \circ \Psi^\varepsilon(t, y_0)|_{\mathbb{C}^D} \leq \tilde{C}_1 e^{-\tilde{C}_2/\varepsilon}. \quad (2.28)$$

4. De plus, si la solution de l'équation (1.1) a une unique solution restant dans K sur $\mathbb{R}$ alors l'équation (2.27) est valable sur $\mathbb{R}$ et la relation (2.28) est valide pour $\tau \in [0, \tilde{T}/\varepsilon^{1+\alpha}]$, pour tout $\tilde{T} > 0$ et $0 < \alpha < 1$ pour ε suffisamment petit.

Préservation des invariants dans le cas d'un système Hamiltonien

Un des avantages de la moyennisation stroboscopique par rapport aux autres choix de moyennisation est la conservation des invariants dans le cas d'équation hamiltonienne. On suppose dans cette partie que D est un nombre pair et que l'équation (1.1) est Hamiltonienne. C'est à dire qu'il existe une fonction $H : \mathbb{T} \times \mathbb{R}^D \rightarrow \mathbb{R}$ analytique par rapport à la seconde variable telle que l'équation (1.1) s'écrive

$$\frac{dy^\varepsilon}{dt} = \varepsilon J^{-1} \nabla H(t, y^\varepsilon), \quad y^\varepsilon(0) = y_0, \quad (2.29)$$

où la matrice J est la matrice symplectique

$$J^{-1} = \begin{pmatrix} 0 & -I_{D/2} \\ I_{D/2} & 0 \end{pmatrix}.$$

On rappelle également qu'une fonction $(t, y) \rightarrow \varphi(t, y)$ est dite symplectique si

$$\forall (t, y, y_1, y_2) \in \mathbb{T} \times K \times (\mathbb{R}^D)^2, \quad (J \partial_y \varphi(t, y) y_1, \partial_y \varphi(t, y) y_2) = (J y_1, y_2)$$

où $(\cdot, \cdot)$ est le produit scalaire usuel sur $\mathbb{R}^D$.

De façon formelle, le champ de vecteur G (de l'équation (1.3)) est Hamiltonien tandis que le changement de variable Φ est une fonction symplectique. Mais, tout comme on ne peut pas avoir l'expression (1.4) de manière exacte, les fonctions $G^{[n]}$ et $\Phi^{[n]}$ définis en (2.8) et (2.9) sont respectivement Hamiltoniennes et symplectiques en commettant une erreur en $\mathcal{O}(\varepsilon^{n+1})$.

Definition 2.8 *Un champ de vecteur $\tilde{f}$ est dit Hamiltonien à une perturbation en ε^{n+1} près si il existe une fonction F telle que*

$$\forall y \in K, \quad \tilde{f}(y) = J^{-1} \nabla F(y) + \mathcal{O}(\varepsilon^{n+1}).$$

De même, une fonction φ est symplectique à une perturbation en ε^{n+1} si elle vérifie pour $(t, y, y_1, y_2) \in \mathbb{T} \times K \times (\mathbb{R}^D)^2$,

$$(J \partial_y \varphi(t, y) y_1, \partial_y \varphi(t, y) y_2) = (J y_1, y_2) + \mathcal{O}(\varepsilon^{n+1} |y_1|_{\mathbb{R}^D} |y_2|_{\mathbb{R}^D}).$$

On a alors le théorème suivant prouvé encore une fois dans [13].

Theorem 2.5 *Sous les hypothèses du Théorème 2.4 et en supposant que f est Hamiltonien (avec la fonction hamiltonienne associée H). Pour tout $n \in \mathbb{N}$, les fonctions $\Phi^{[n]}$ et $G^{[n]}$ définis par (2.24) et (2.25) sont respectivement symplectique*

et Hamiltoniens a un terme de perturbation en ε^{n+1} . C'est à dire que pour $(t, y) \in \mathbb{T} \times K, y_1, y_2 \in \mathbb{R}^D$, on a

$$(J\partial_y \Phi^{[n]}(t, y)y_1, \partial_y \Phi^{[n]}(t, y)y_2) = (Jy_1, y_2) + \mathcal{O}(\varepsilon^{n+1}|y_1|_{\mathbb{R}^D}|y_2|_{\mathbb{R}^D})$$

$$G^{[n]}(y) = J^{-1}\nabla H^{[n]}(y) + \mathcal{O}(\varepsilon^{n+1}),$$

où la fonction $H^{[n]}$ est défini par

$$H^{[n]}(y) = \langle H \circ \Phi^{[n+1]}(\cdot, y) \rangle - \frac{1}{2\varepsilon} \left\langle (J\partial_t \Phi^{[n+1]}(\cdot, y), \Phi^{[n+1]}(\cdot, y)) \right\rangle.$$

On sait alors que la procédure de moyennisation stroboscopique permet de conserver le caractère Hamiltonien de l'équation de base. En fait, il préserve tout invariant du système (1.1). Plus précisément, on suppose que la solution y^ε de (1.1) admette un invariant Q , c'est à dire une fonction régulière de $\mathbb{T} \times \mathbb{R}^D \rightarrow \mathbb{R}$ telle que

$$\forall t \in [0, T/\varepsilon], \quad Q(t, y^\varepsilon(t)) = Q(0, y_0).$$

On remarque qu'en dérivant les deux côtés de la dernière équation, on obtient

$$\partial_t Q(t, y^\varepsilon) + \varepsilon Q(t, y^\varepsilon) f(t, y^\varepsilon) = 0. \quad (2.30)$$

On a alors le théorème suivant relatif à la préservation des invariants.

Theorem 2.6 *On se place sous les hypothèses du Théorème 2.4, et on suppose que la fonction $(t, y) \in \mathbb{T} \times \mathbb{R}^D \rightarrow Q(t, y) \in \mathbb{R}$ est un invariant de l'équation (1.1) dans le sens que Q satisfait l'équation (2.30) pour tout $(t, y) \in \mathbb{T} \times K$. On suppose de plus que Q est analytique par rapport à la variable y sur K_ρ pour $0 < \rho < R$. Alors, pour tout $n \in \mathbb{N}$, le changement de variable $\Phi^{[n]}$ et le champ de vecteur moyenné $G^{[n]}$ satisfont pour tout $y \in K, t \in \mathbb{T}$,*

$$Q(t, \Phi^{[n]}(t, y)) = Q(0, y) + \mathcal{O}(\varepsilon^{n+1}), \quad \text{et} \quad (\partial_y Q(0, y))G^{[n]}(y) = \mathcal{O}(\varepsilon^n).$$

Remark 2.7 *La preuve du théorème 2.6 présentée dans [13] utilise l'hypothèse que l'invariant Q est analytique. Afin d'obtenir des erreurs, en ε^{n+1} , peut conjecturer qu'il suffit que Q soit simplement de classe $\mathcal{C}^n$ par rapport à la variable y .*

3 Une méthode numérique: La SAM (Stroboscopic Averaging Method)

La difficulté de calculer numériquement les solutions des équations (1.1) ou (1.2) avec des méthodes standard (Runge Kutta, Méthodes de Taylor, ...) est bien connue. Cela provient notamment du fait que le pas d'intégration doit être un $\mathcal{O}(1)$

alors que le temps d'intégration est en T/ε (ou bien le pas d'intégration doit être choisi inférieur à ε pour des problèmes hautement oscillants) ce qui augmente de manière drastique les temps de calculs. L'idée à la base de la moyennisation étant de substituer l'équation oscillante par l'équation moyennée autonome (1.3)

$$\frac{d}{dt}\Psi(t, y_0) = \varepsilon G(\Psi(t, y_0)),$$

on cherche des méthodes numériques qui approchent le flot $\Psi(t, y_0)$.

Malheureusement, le champ de vecteur G est très compliqué à calculer numériquement (que ce soit en utilisant les formules de Taylor, les B-séries ou l'équation de transport). Même en ne s'intéressant qu'au premier termes du développement de G , c'est à dire aux fonctions G_1 et G_2 dans la formule (2.8), les calculs numériques restent importants.

3.1 Principe général de la SAM

La *Stroboscopic Averaging Method* (ou SAM) développée par Chartier & al. dans l'article [9] s'inspire des méthodes dites multi-échelles afin d'intégrer numériquement l'équation moyennée autonome (2.14) sans jamais calculer le champ de vecteur G . Elle est basée sur la relation (formelle)

$$y^\varepsilon(t) = \Phi(t, \Psi(t, y_0))$$

qui donne directement

$$\forall \mathbf{t} \in \mathbf{T}_S, \quad y(\mathbf{t}) = \Psi(\mathbf{t}, y_0), \quad (3.1)$$

où l'on rappelle que $\mathbf{T}_S = \{Pk\varepsilon, k \in \mathbb{Z}\}$ est l'ensemble des temps dit stroboscopiques. Bien que les égalités écrites précédemment soient formelles, elles peuvent être rendues rigoureuses en considérant les versions tronquées exposées précédemment:

$$\left| y(t) - \Phi^{[n]}(t, \Psi^{[n]}(t, y_0)) \right| = \mathcal{O}(\varepsilon^n), \quad \forall t \in [0, T/\varepsilon].$$

Cependant, n pouvant être choisi aussi grand que l'on veut (sous certaines hypothèses sur la régularité de f , voir partie 2), l'erreur commise en $\mathcal{O}(\varepsilon^n)$ est négligeable en comparaison des autres erreurs qui interviennent dans la méthode numérique.

Dans l'esprit des méthodes multi-échelles, on choisit une méthode macro ayant un pas d'intégration Δt supérieur à la période P de la fonction f afin d'intégrer numériquement l'équation

$$\frac{d\Psi}{dt}(t, y_0) = \varepsilon G(\Psi(t, y_0)). \quad (3.2)$$

Si l'on utilise par exemple des méthodes de Runge Kutta pour la méthode macro, les seules informations dont on a besoin sont les valeurs de $G(y^*)$ pour $y^* \in \mathbb{R}^D$. L'information nécessaire pour calculer G est alors donnée par une micro-méthode ayant un pas d'intégration h suffisamment petit pour pouvoir intégrer numériquement l'équation (1.1). (Les méthodes macro et micro n'étant pas forcément les même).

Par définition, on a

$$G(y^*) = \frac{1}{\varepsilon} \frac{d\Psi}{dt}(t, y^*) \Big|_{t=0} = \frac{d\Psi}{d\tau} \left(\frac{\tau}{\varepsilon}, y^* \right) \Big|_{\tau=0}$$

que l'on peut approcher par une formule des différences finies centrées

$$G^\varepsilon(y^*) = \frac{1}{2\delta} \left(\Psi \left(\frac{\delta}{\varepsilon}, y^* \right) - \Psi \left(-\frac{\delta}{\varepsilon}, y^* \right) \right) + \mathcal{O}(\delta^2).$$

En notant α^ε le flot de l'équation de départ (1.1), en prenant $\delta = P\varepsilon$ et en utilisant la relation (3.1), on obtient

$$G^\varepsilon(y^*) = \frac{1}{2P\varepsilon} (\alpha^\varepsilon(P, y^*) - \alpha^\varepsilon(-P, y^*)) + \mathcal{O}(\varepsilon^2). \quad (3.3)$$

Les quantités $\alpha^\varepsilon(P, y^*)$ et $\alpha^\varepsilon(-P, y^*)$ peuvent alors être calculés numériquement en intégrant l'équation (1.1) (en prenant pour condition initiale y^*) avec la méthode micro entre 0 et P puis entre 0 et $-P$.

On peut noter que l'approximation par différences finies qui conduit à l'expression (3.3) peut être modifiée afin d'obtenir des erreurs en $\mathcal{O}(\varepsilon^3)$, $\mathcal{O}(\varepsilon^4)$... Par exemple, on a

$$G^\varepsilon(y^*) = \frac{1}{12P\varepsilon} \left(-\alpha^\varepsilon(2P, y^*) + 8\alpha^\varepsilon(P, y^*) - 8\alpha^\varepsilon(-P, y^*) + \alpha^\varepsilon(-2P, y_0) \right) + \mathcal{O}(\varepsilon^4). \quad (3.4)$$

3.2 Estimations d'erreurs et exemples numériques

On analyse dans cette partie l'erreur numérique commise par la SAM. On se souvient que la méthode consiste à intégrer (3.2) avec une macr- méthode sachant que G est donné par (3.3) et que le flot α^ε est approché par une micro-méthode. Il y a donc trois considérations à prendre en compte dans l'erreur.

- L'erreur commise par l'approximation due à la macro-méthode .
- L'erreur induite par l'approximation de G^ε par une méthode de différence finie (cf formules (3.3) et (3.4)).
- L'erreur commise lors de l'approximation du flot α^ε par la micro-méthode.

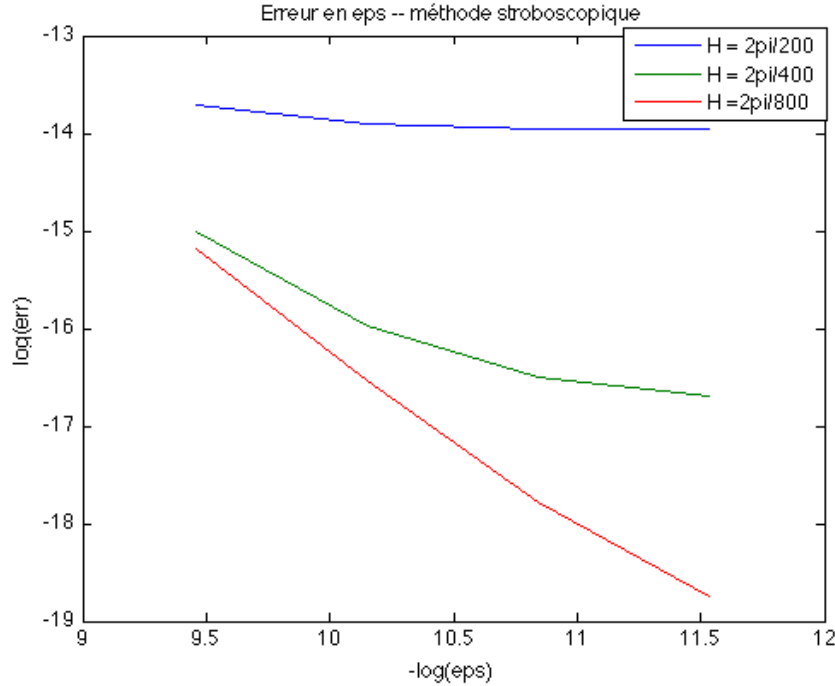


Figure 1.5: Erreur numérique de la SAM en ε pour différent pas d'intégration Δt .

On donne maintenant précisément les erreurs commises dans chacun des cas. Tout d'abord, si l'on considère que la macro-méthode est d'ordre p et a un pas d'intégration Δt , l'erreur commise sera en $\mathcal{O}((\Delta t)^p)$.

L'erreur commise par l'approximation de la fonction G^ε dépend de la méthode des différences finies utilisée. Néanmoins l'erreur sera en $\mathcal{O}(\varepsilon^\delta)$ avec $\delta = 2$ dans l'expression (3.3) et $\delta = 4$ dans l'expression (3.4).

Enfin, l'erreur due à l'approximation du flot α^ε par une micro-méthode d'ordre q et de pas d'intégration h est bornée par $\frac{1}{\varepsilon}h^q$. Finalement l'erreur totale de la méthode de moyennisation stroboscopique est

$$\mathcal{O}\left((\Delta t)^p + \varepsilon^\delta + \frac{1}{\varepsilon}h^q\right).$$

D'autres remarques sur l'erreur produite par la SAM ainsi que de nombreux exemples numériques peuvent être trouvés dans [9].

Afin d'étayer l'affirmation précédente, on présente un exemple des erreurs en ε (Figure 1.5) et en Δt (Figure 1.6). On a ainsi calculer numériquement par la SAM, les solutions de l'équation (1.11) (problème jouet) que l'on a confronté à la solution

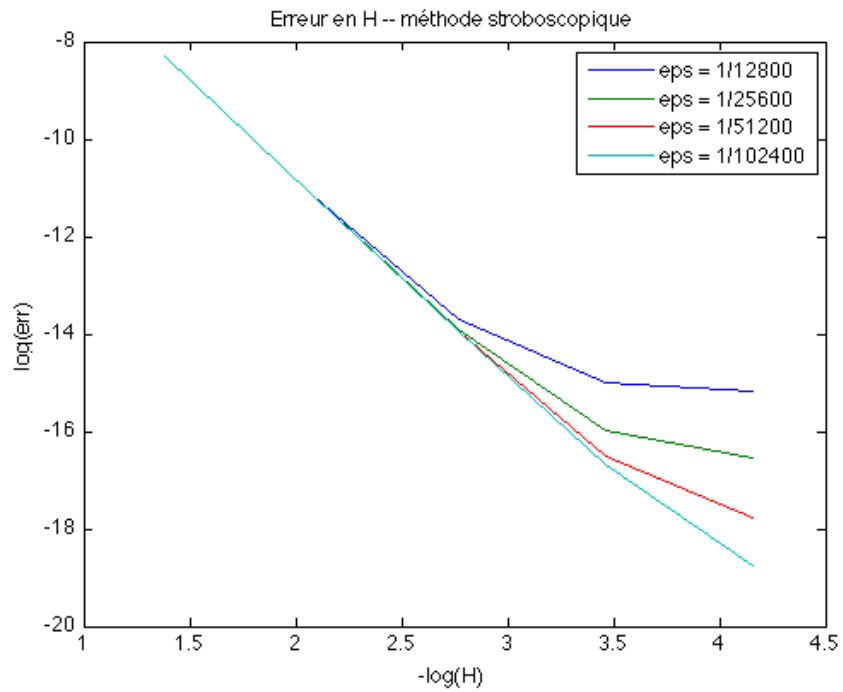


Figure 1.6: Erreur numérique de la SAM en Δt pour différentes valeurs de ϵ .

exacte du problème. La condition initiale est $y_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. On a appliqué une méthode de RK4 pour la macro-méthode et la micro-méthode (avec un pas d'intégration pour la micro méthode suffisamment petit pour que l'erreur due à la micro-méthode soit négligeable) tandis que l'on a utilisé la formule (3.3) pour calculer G . Les résultats obtenus sont ceux espérés avec un phénomène de saturation qui s'explique par la présence des autres termes présents dans l'erreur.

4 Les problèmes quasi-périodiques

Comme annoncé dans la partie 1.2, on s'intéresse en second lieu aux problèmes quasi-périodique. Pour $d \geq 2$, on considère les équations de la forme (1.9),

$$\frac{dy^\varepsilon}{dt} = \varepsilon f(t\omega_1, \dots, t\omega_d, y^\varepsilon), \quad y(0) = y_0. \quad (4.1)$$

On supposera que la fonction $f : \mathbb{T}^d \times X \rightarrow X$ est régulière. Dans cette équation, le vecteur $\omega = (\omega_1, \dots, \omega_d)$ est appelé vecteur des fréquence de f . Pour simplifier les notations, on écrira $f(t\omega, y) = f(t\omega_1, \dots, t\omega_d, y)$. Afin de gérer ce qu'on appelle les petits diviseurs ([42, 32]) le vecteur de fréquence ω doit satisfaire une condition de non résonnance forte (ou condition diophantine de Siegel) définie ci-après par l'hypothèse 1.

Assumption 1 *Il existe deux constantes positives γ et ν , telles que pour tout k non nul dans $\mathbb{Z}^d$, la relation suivante soit vraie.*

$$|k \cdot \omega| \geq \gamma |k|^{-\nu},$$

où l'on note pour tout éléments de $\mathbb{Z}^d$, $|k| = \sum_{i=1}^d |k_i|$.

L'hypothèse ci-dessus semble restreindre drastiquement le choix des vecteurs ω mais il n'en est rien. En effet, il a été montré (cf [42]) que pour $\nu > d$, l'ensemble des vecteurs ω qui ne vérifient pas l'hypothèse 1 est de mesure nulle dans $\mathbb{R}^d$.

On cherche alors, comme dans le cas mono-fréquence (cf partie 2) un changement de variable quasi-périodique $(\theta, y) \in \mathbb{T}^d \times X \rightarrow \Phi(\theta, y)$ ainsi que la solution Ψ d'une équation autonome

$$\frac{d\Psi}{dt}(t, y_0) = \varepsilon G^\varepsilon(\Psi(t, y_0))$$

tel que l'on ait

$$y^\varepsilon(t) = \Phi(t\omega, \Psi(t, y_0)). \quad (4.2)$$

On dérive formellement l'égalité (4.2) par rapport au temps, ce qui donne en notant $\omega \cdot \theta = \sum_{j=1}^d \omega_j \partial_{\theta_j}$

$$\frac{dy}{dt}(t) = (\omega \cdot \partial_\theta) \Phi(\theta, \Psi(t, y_0))|_{\theta=t\omega} + \varepsilon \partial_y \Phi(t\omega, \Psi(t, y_0)) G^\varepsilon(\Psi(t, y_0))$$

D'un autre côté, on a

$$\frac{dy^\varepsilon}{dt}(t) = \varepsilon f(t\omega, \Phi(t\omega, \Psi(t, y_0)))$$

En prenant $y = \Psi(t, y_0)$ et en remarquant que l'espace $\{t\omega, t \in \mathbb{R}\}$ est dense dans $\mathbb{T}^d$, on trouve l'équation homologique pour $(\theta, y) \in \mathbb{T}^d \times X$,

$$(\omega \cdot \partial_\theta) \Phi(\theta, y) = \varepsilon (f \circ \Phi(\theta, y) - \partial_y \Phi(\theta, y) G(y)). \quad (4.3)$$

Une fois l'équation homologique écrite, les méthodes sont les mêmes que dans le cas périodique. Ainsi Perko (cf [44]), donne un théorème de moyennisation pour un champ de vecteur quasi-périodique. C'est également le cas de Chartier & al. (cf [17, 18]) qui présentent un théorème de moyennisation en utilisant les B-séries. Enfin, le chapitre 3 présente un théorème de moyennisation dans le cadre plus général où X est un espace de Hilbert séparable de dimension infinie. On présente dans cette partie les travaux de Chartier & al.

On suppose que l'équation 4.1 est étudiée sur un domaine $\mathcal{K} \subset \mathbb{R}^D$. On considère alors $\|\cdot\|$ la norme sur $\mathbb{C}^D$ et l'ensemble $\mathcal{K}_\rho \subset \mathbb{C}^D$ donné par

$$\mathcal{K}_\rho = \{y + \tilde{y} \in \mathbb{C}^D; y \in \bar{\mathcal{K}}, \|z\| \leq \rho\}.$$

Finalement, on définit la norme sur les fonctions bornées ϕ défini sur $\mathcal{K}_\rho$ par

$$\|\phi\|_\rho = \sup_{y \in \mathcal{K}_\rho} \|\phi(y)\|.$$

L'hypothèse de travail pour la fonction f de l'équation (4.1) est alors donnée par

Assumption 2 *Il existe $R > 0$, $\mu > 0$, et un ouvert $\mathcal{U} \subset \mathcal{K}_\rho$ telle que pour tout $\theta \in \mathbb{T}^d$, La fonction $f(\cdot, \theta)$ est prolongeable en une fonction analytique de $\mathcal{U}$. De plus, on suppose que les coefficients de Fourier f_k de la fonction f satisfassent*

$$\forall k \in \mathbb{Z}^d, \quad \|f_k\|_R \leq a_k e^{-\mu|k|}$$

avec $M := \sum_{k \in \mathbb{Z}^d} a_k < \infty$.

Avec ces hypothèse, on peut prouver le théorème suivant.

Theorem 4.1 *On suppose que f satisfait l'hypothèse 2 et que $\omega \in \mathbb{R}^d$ ($d > 1$) satisfait l'hypothèse 1. Il existe alors $\varepsilon_0(n) > 0$ tel que pour tout $|\varepsilon| \leq \varepsilon_0(n)$, la solution y^ε de l'équation (4.1) satisfait*

$$\forall t \in [0, T/\varepsilon], \quad y^\varepsilon(t) = \Phi^\varepsilon(t\omega, Y^{[n]}(t))$$

où $Y^{[n]}$ est la solution de l'équation quasi-autonome suivante

$$\frac{dY^{[n]}}{dt} = \varepsilon G^{[n]}(Y) + \mathcal{R}^{[n]}(t\omega, Y). \quad (4.4)$$

Le reste $\mathcal{R}^{[n]}$ ci-dessus vérifie

$$\|\mathcal{R}^{[n]}\|_R \leq \frac{5|\varepsilon/\varepsilon_0|}{1 - |\varepsilon/\varepsilon_0|} M.$$

De plus, il existe $\varepsilon^* > 0$ tel que pour $|\varepsilon| \leq \varepsilon^*$, il existe $n(\varepsilon)$ tel que

$$\|\mathcal{R}^{[n(\varepsilon)]}\|_{R/2} \leq C_1 \exp\left(-\frac{C_2}{|\varepsilon|^{1/(\nu+1)}}\right),$$

où C_1 et C_2 sont deux constantes positives indépendantes de ε .

5 Résumé des résultats de la thèse

Cette dernière partie de l'introduction reprend tous les résultats de la thèse chapitre par chapitre et explique en quoi ces résultats sont nouveaux dans le contexte de la moyennisation stroboscopique.

5.1 Différentes méthodes pour différents point de vue

Même si toutes les méthodes présentées dans la partie 2 amènent aux mêmes résultats, les méthodes diffèrent par leur mise en oeuvre et leur philosophie. Ainsi la méthode par séries de Taylor est plus intuitive. Les formules obtenues sont compactes et facilement compréhensibles puisqu'elles ne dépendent que de la fonction f de l'équation de départ (1.1)(et de ses dérivées). Cependant la mise en évidence de la préservation des invariants et de la géométrie du problème par cette méthode n'est pas évidente. On présente par exemple dans le chapitre 2 les Hamiltoniens des deux premiers termes du champs de vecteurs moyennée (les fonctions G_1 et G_2 de la formule (2.8)). Il n'est pas facile de trouver par ce biais les Hamiltoniens pour des termes plus élevés ($G_3, G_4, \dots$).

A l'inverse, la méthode utilisant les B-séries qui consiste à décomposer la fonction f en séries de Fourier, permettent de montrer facilement la préservation des invariants et de trouver les Hamiltoniens associés au champs de vecteur moyenné (cf [16]). Par contre, les résultats sont donnés sous formes de sommes des modes de Fourier de la fonction f . Ces formules sont du coup plus difficiles à appréhender et demandent un investissement pour être comprise.

Entre ces deux extrêmes, se situe la méthode par équation de transport. Les formules sont moins éclatées que dans le cas des B-séries puisqu'on n'a pas besoin de décomposer la fonction f en séries de Fourier. Cependant, les formules reposent sur la résolution (approchée) d'une équation de transport et ne sont donc pas aussi explicite que dans le cas de la méthode de Taylor. L'avantage principal de cette méthode est que l'on peut montrer facilement la préservation des invariants (cf [13]) sans passer par les B-séries.

Dans la table 1.2, on donne les différentes contributions dans le domaine de la moyennisation stroboscopique d'ordre élevée. On doit faire quelques remarques sur ce qu'il contient. Ainsi, les travaux de Perko sont exposés dans un cadre qui n'est pas celui de la moyennisation stroboscopique mais uniquement la moyennisation usuelle (voir section 1.2). Le cas quasi-périodique en dimension infinie est traité dans le chapitre 3. Cependant, aucun exemple de dimension infinie ne vient étayer l'analyse. Il faut user d'un subterfuge présenté dans le chapitre 4 pour appliquer

Table 1.2: Résumé des différents travaux dans le cadre de la moyennisation stroboscopique d'ordre élevée.

	Méthode des séries de Taylor	Méthode des B-séries	Méthode avec équation de transport
Cas périodique, dimension finie	Perko [44] (Cas polynomial)	Chartier & al. [16] (Cas analytique)	Castella & al. [13] (Cas analytique)
Cas quasi-périodique, dimension finie	Perko [44] (Cas polynomial)	Chartier & al. [18] (Cas analytique)	Chapitre 3 (Cas analytique)
Cas périodique, dimension infinie	Chapitre 2 (Cas analytique)		Castella & al. [13] (Cas analytique)
Cas quasi-périodique, dimension infinie			Chapitre 3 et 4 (Cas polynomial)

la procédure de moyennisation quasi-stroboscopique à une équation de Schrödinger non linéaire.

5.2 Chapitre 2 : Méthode de moyennisation stroboscopique et équation des ondes hautement oscillantes

Dans ce chapitre, on considère la méthode de moyennisation stroboscopique appliquée à des équations d'évolutions hautement oscillantes, posées dans un espace de Banach X , de la forme

$$\frac{\partial}{\partial t}u(t) = g\left(\frac{t}{\varepsilon}, u, \varepsilon\right), \quad u(0) = u_0 \in X, \quad t \in \mathbb{R}, \quad \varepsilon > 0. \quad (5.1)$$

La fonction g est supposée P -périodique par rapport à sa première variable, régulière par rapport à la seconde variable et au moins continue par rapport à la troisième variable. Ce problème est similaire au problème hautement oscillant (1.2) dans un espace de Banach de dimension infinie.

On étend ici le travail de Perko (cf [44] et section 2) dans trois directions: On adapte la construction de Perko à la moyennisation stroboscopique, on donne les résultats pour des équations dans des espaces de Banach de dimension infinie et on va au delà des erreurs polynomiales pour montrer des estimations d'erreurs exponentiellement petites (sous réserve que la fonction g soit analytique).

Il s'agit dans un premier temps, de redéfinir l'analyticité des fonctions dans un espace de Banach. Ainsi, on introduit la complexification de l'espace X

$$X_{\mathbb{C}} = \{U := u + i\tilde{u}, (u, \tilde{u}) \in X^2\}.$$

L'espace $X_{\mathbb{C}}$ est un espace de Banach sous réserve de lui associer la norme

$$\|U\|_{X_{\mathbb{C}}} := \sup_{z \in \mathbb{C}^*} \frac{\|\Re(zU)\|_X}{|z|}.$$

De plus pour tout $\rho > 0$ et K un ouvert de X , on considère le plongement ouvert de K de taille $\rho > 0$ dans l'espace $X_{\mathbb{C}}$ par

$$K_{\rho} = \{u + \tilde{u} : (u, \tilde{u}) \in O \times X_{\mathbb{C}}, \|\tilde{u}\|_{X_{\mathbb{C}}} < \rho\}.$$

Nous sommes alors prêt à donner les hypothèses sur g .

Assumption 3 *La fonction g est réelle analytique, c'est à dire qu'il existe $R > 0$ et $\varepsilon^* > 0$ tel que g soit une fonction analytique et bornée par M sur $\mathbb{T} \times K_{2R} \times [0, \varepsilon^*]$.*

Le théorème principal de ce chapitre est alors le théorème de moyennisation stroboscopique avec estimation d'erreur exponentiellement petite.

Theorem 5.1 *Soit $u_0 \in X$ et g satisfaisant l'hypothèse 3. Alors il existe $\varepsilon_0 > 0$ tel que pour tout $0 < \varepsilon \leq \varepsilon_0$, il existe $N(\varepsilon) \in \mathbb{N}$ tel que*

1. *La solution de l'équation (5.1) et l'approximation $u_{\varepsilon} := u_{N(\varepsilon)}$ défini par la formule (2.10) existent et sont inclus respectivement dans $O_{R/2}$ et O_R pour $t \in [0, T_f]$.*

2. *On a*

$$\sup_{t \in [0, T_f]} \|u(t) - u_{\varepsilon}(t)\|_{X_{\mathbb{C}}} \leq C_1 \exp\left(-\frac{C_2}{\varepsilon}\right)$$

où C_1 et C_2 sont deux constantes positives indépendantes de ε .

Dans le cas où l'espace X est un espace de Hilbert et que le champ de vecteur initial g est Hamiltonien, il a été montré que les champs de vecteurs moyennés G_j (formule (2.8)) sont également des champs de vecteurs Hamiltoniens. Dans ce chapitre, on exhibe les Hamiltoniens correspondant aux champs de vecteurs G_1 et G_2 .

Dans la partie 3, on applique le théorème (5.1) sur deux exemples. Le premier exemple est l'équation des ondes semi-linéaire et hautement oscillante

$$\frac{\partial^2 u}{\partial t^2} - \frac{1}{\varepsilon^2} \frac{\partial u}{\partial x^2} = \frac{1}{\varepsilon} f(u), \quad t \in \mathbb{R}, \quad x \in [0, 1] \quad (5.2)$$

avec des conditions de Dirichlet au bord du domaine. Le second exemple est l'équation des ondes semi-linéaire

$$\frac{\partial^2 u}{\partial t^2} - \frac{1}{\varepsilon^2} \frac{\partial^2 u}{\partial x^2} = f(u), \quad t \in \mathbb{R}, \quad x \in [0, 1]$$

muni de conditions périodiques au bord du domaine.

Enfin, la dernière partie du chapitre est consacrée à l'application de la SAM à l'équation des ondes (5.2) (avec condition de Dirichlet au bord du domaine). L'utilisation de la SAM pour des équations d'évolutions est alors novatrice. Elle a dans le même temps et par la suite été testée sur d'autres équations d'évolutions (des équations de Schrödinger dans [13, 19]. Après avoir vérifié que les estimations d'erreurs de la méthode numérique étaient semblables au cas des équations différentielles ordinaires (cf section 3), on s'intéresse à la dynamique des modes de Fourier de l'équation des ondes. L'étude de la dynamique de l'équation moyennée permet de mettre en évidence, heuristiquement, des phénomènes périodiques ou presque-périodiques.

5.3 Chapitre 3 : Méthode de moyennisation quasi-stroboscopique dans des espaces de Hilbert

Dans ce chapitre, on s'intéresse aux équations d'évolutions posées dans un espace de Hilbert réel X de la forme

$$\frac{d}{dt}u^\varepsilon(t) = \varepsilon g(t, u^\varepsilon(t)), \quad u^\varepsilon(0) = u_0 \in X, \quad t \in [0, T/\varepsilon] \quad (5.3)$$

avec $g(t, u) = G(t\omega_1, \dots, t\omega_d, u)$ et $\omega = (\omega_1, \dots, \omega_d) \in \mathbb{R}^d$. La fonction G est une fonction de $\mathbb{T}^d \times X \rightarrow X$ et est supposée 1-périodique par rapport à chaque variable de $\mathbb{T}^d$, régulière par rapport à $\theta \in \mathbb{T}^d$ et $u \in X$. La fonction g est ainsi quasi-périodique et le problème est similaire à ceux de la partie 4 dans des espaces de dimension infinie. Le vecteur des fréquences ω devra donc satisfaire une condition de non résonance forte (hypothèse 1). L'étude est alors faite dans le cas où la fonction g est à la fois analytique par rapport à $\theta \in \mathbb{T}^d$ et par rapport à $u \in X$. Sous cette condition, on peut écrire un théorème similaire au Théorème 5.1.

La première partie du chapitre est donc consacrée à la présentation des fonctions $f : \mathbb{T}^d \times X \rightarrow X$ quasi-périodique et analytique par rapport aux deux variables. On considère un ouvert K et on définit alors, pour $\rho > 0$, $\mathbf{A}_\rho$ l'ensemble des fonctions de X dans X analytique et bornée dans K_ρ . La définition de K_ρ étant la même que dans la partie précédente (section 5.2). On définit ensuite l'ensemble $\mathcal{A}_\rho^\omega$ des fonctions $f \in \mathcal{C}^{d+1}(\mathbb{R}, \mathbf{A}_\rho)$ quasi-périodiques de vecteur de fréquence ω , c'est à dire qu'il existe une fonction $F \in \mathcal{C}^{d+1}(\mathbb{T}^d, \mathbf{A}_\rho)$ telle que

$$\forall t \in \mathbb{R}, \forall u \in X, \quad f(t, u) = F(t\omega_1, \dots, t\omega_d, u).$$

Enfin, on définit l'ensemble $\mathcal{A}_{\sigma, \rho}^\omega$ des fonctions $f \in \mathcal{A}_\rho^\omega$ qui sont analytiques sur

$$D_\sigma = \{\theta = \theta_1 + i\theta_2, (\theta_1, \theta_2) \in \mathbb{T}^d \times \mathbb{R}^d, \|\theta_2\|_\infty < \sigma\}$$

et pour lesquelles la fonction F correspondante est bornée. Cela nous permet d'écrire la condition sur g :

Assumption 4 *Il existe R, W et C trois constantes positives telles que la fonction g appartienne à $\mathcal{A}_{2W,2R}^\omega$ et soit bornée par C .*

La deuxième partie du chapitre est alors consacré à la preuve du théorème suivant.

Theorem 5.2 *On suppose que les hypothèses 1 et 4 sont satisfaites. On note $\eta = (\nu + d + 1)^{-1}$. Alors, il existe $\hat{\varepsilon} > 0$ tel que pour tout $0 < \varepsilon \leq \hat{\varepsilon}$, on construise un changement de variable $\phi^\varepsilon \in \mathcal{A}_{W,R}^\omega$ et un champ de vecteur autonome $\mathcal{G}^\varepsilon \in \mathbf{A}_R$ satisfaisant*

$$\forall t \in [0, T/\varepsilon], \quad \|u^\varepsilon(t) - \phi^\varepsilon(t, \psi^\varepsilon(t, u_0))\|_X \leq C_1 \exp(-C_2 \varepsilon^{-\eta}),$$

où C_1 et C_2 sont deux constantes positives indépendantes de ε et où ψ^ε est le flot de l'équation autonome $\frac{d}{dt}\psi^\varepsilon(t, u_0) = \varepsilon \mathcal{G}^\varepsilon(\psi^\varepsilon(t, u_0))$.

5.4 Chapitre 4 : Méthode de moyennisation quasi-stroboscopique pour une équation de Schrödinger semi-linéaire.

Dans ce chapitre, on cherche à appliquer le théorème 5.2 du chapitre précédent à l'équation de Schrödinger non linéaire suivante:

$$\begin{aligned} i\partial_t y^\varepsilon(t, x) &= - \sum_{j=1}^d \omega_j \frac{\partial^2}{\partial x_j^2} y^\varepsilon(t, x) + \varepsilon f(|y^\varepsilon(t, x)|^2) y^\varepsilon(t, x), \quad t \geq 0, x \in \mathbb{T}^d \\ y^\varepsilon(0, \cdot) &= y_0. \end{aligned} \tag{5.4}$$

Dans cette équation d est un entier supérieur à 2, f est une fonction analytique réelle et $\omega = (\omega_1, \dots, \omega_d)$ satisfait l'hypothèse 1. Le problème principal est que l'équation (5.4) ne peut pas être mis sous la forme (5.3) avec une fonction g analytique par rapport à θ (cela vient notamment du fait que l'opérateur $t \rightarrow e^{t\omega_j \frac{\partial^2}{\partial x_j^2}}$ n'est pas analytique par rapport à t).

La première chose que l'on fait dans ce chapitre est de transformer l'équation 5.4 afin de la réécrire sans les nombres complexes. On note alors $\Re(y)$ et $\Im(y)$ la partie réelle et imaginaire de $y \in \mathbb{C}$ et on définit le changement de variable

$$u^\varepsilon(t, x) = \begin{pmatrix} \Im(\tilde{y}^\varepsilon(t, x)) \\ \Re(\tilde{y}^\varepsilon(t, x)) \end{pmatrix} \in \mathbb{R} \times \mathbb{R}.$$

Pour $s > \frac{d}{2}$, on travaille alors dans les espaces

$$X = H_{\mathbb{R}}^s(\mathbb{T}^D) \times H_{\mathbb{R}}^s(\mathbb{T}^D) \quad \text{and} \quad Z = L_{\mathbb{R}}^2(\mathbb{T}^D) \times L_{\mathbb{R}}^2(\mathbb{T}^D)$$

où l'espace H^s est l'espace de Sobolev des fonctions s -dérivables au sens des distributions et dont toutes les dérivées sont dans L^2 . Enfin, on note pour $j = 1, \dots, d$,

$$A_j = \begin{pmatrix} 0 & \frac{\partial^2}{\partial x_j^2} \\ -\frac{\partial^2}{\partial x_j^2} & 0 \end{pmatrix} \in \mathcal{L}(X, Z) \quad \text{et} \quad \omega \cdot A = \sum_{j=1}^d \omega_j A_j.$$

Finalement, écrivant $u = \begin{pmatrix} p \\ q \end{pmatrix} \in X$, $|u|_2^2 = p^2 + q^2$ l'équation (5.4) est équivalente à

$$\begin{aligned} \partial_t u^\varepsilon(t, x) &= (\omega \cdot A) u^\varepsilon(t, x) + \varepsilon h(u^\varepsilon(t, x)), \quad t \geq 0, x \in \mathbb{T}^d \\ u^\varepsilon(0, \cdot) &= u_0 := \begin{pmatrix} \mathfrak{I}(y_0) \\ \mathfrak{R}(y_0) \end{pmatrix} \in X, \end{aligned} \tag{5.5}$$

avec

$$h : u \in X \rightarrow X \ni f(|u|_2^2) J^{-1} u, \quad \text{and} \quad J^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Notre but est alors d'appliquer le théorème 5.2 sur un sous espace de X de dimension fini. On introduit alors pour tout $L \in \mathbb{N}^*$, l'espace

$$Z_L = \text{Vect} \left\{ x \rightarrow e^{2i\pi(k \cdot x)}, \quad k \in \mathbb{Z}^D \text{ s.t. } |k| \leq L \right\} \subset Z.$$

On note alors X_L l'espace vectoriel de dimension finie $X \cap Z_L$ et on introduit la projection

$$\Pi_L : u \in X \rightarrow X_L \ni (t, x) \rightarrow \sum_{|k| \leq L} \hat{u}_k e^{2i\pi(k \cdot x)},$$

où les $\hat{u}_k$ sont les coefficients de Fourier de la fonction u . Cet artifice nous permet alors de prouver le théorème (on rappelle que la constante ν est définie dans l'hypothèse 1).

Theorem 5.3 *Soit $s > \max(d/2, 4(\nu + d))$ et $n \leq \frac{s}{4(\nu + d)}$. Alors, pour ε suffisamment petit, il existe un changement de variable Φ^ε quasi-périodique et analytique ainsi qu'un champ de vecteur $\mathcal{G}^\varepsilon$ tel que*

$$\forall t \in [0, T/\varepsilon], \quad \|u^\varepsilon(t) - \Phi^\varepsilon(t, \psi^\varepsilon(t, \Pi_L u_0))\|_Z \leq C\varepsilon^n$$

où u^ε est la solution de l'équation (5.5), ψ^ε est la solution de l'équation autonome

$$\frac{d}{dt} \psi^\varepsilon(t, \Pi_L u_0) = \varepsilon \mathcal{G}^\varepsilon(\psi^\varepsilon(t, \Pi_L u_0))$$

et où C est une constante positive indépendante de ε .

Chapter 2

Stroboscopic averaging of highly-oscillatory non-linear wave equations

1 Introduction

In this article, we consider the so-called *stroboscopic* averaging technique applied to general, highly-oscillatory evolution equations, posed in a Banach space X , of the form

$$\frac{d}{dt}u(t) = g\left(\frac{t}{\varepsilon}, u, \varepsilon\right), \quad u(0) = u_0 \in X, \quad t \in \mathbb{R}, \quad \varepsilon > 0. \quad (1.1)$$

Here g is assumed to be T -periodic with respect to its first variable $\tau = \frac{t}{\varepsilon}$ (we shall denote accordingly $\mathbb{T} = \mathbb{R}/T\mathbb{Z}$), smooth in $u \in X$ and continuous in the third variable. When ε is small, this system is highly-oscillatory with a single frequency. In a seminal paper [44], Perko constructed the averaged equations at arbitrary order in ε for highly-oscillatory ordinary differential equations (ODEs), and obtained error estimates. Here, we extend this analysis in three directions: we adapt Perko's construction to the stroboscopic averaging, we set up the equations in the infinite dimensional context and we go beyond polynomial errors to produce an exponential error estimate. In contrast with Castella et al. [13], where a similar problem is studied, our analysis is based on formulas derived in [44], and leads to explicit constants in both polynomial and exponential error estimates.

Generally speaking, the goal of stroboscopic averaging is to find an averaged system of the form

$$\begin{aligned} \frac{dV}{dt} &= G(V; \varepsilon) := G_1(V) + \varepsilon G_2(V) + \dots, \quad t \in \mathbb{R}, \\ V(0) &= u_0. \end{aligned} \quad (1.2)$$

whose solution approximates the solution of system (1.1) at times $t_n = nT\varepsilon$ (called "stroboscopic times") that are multiples of the t -period.

In accordance with [44], the first order averaged model

$$\frac{d}{dt}V_1 = G_1(V_1), \quad \text{with} \quad G_1(V_1) := \frac{1}{T} \int_0^T g(\tau, V_1, \varepsilon) d\tau,$$

will play a specific role. We will assume that this system has a unique solution in an open ball $B \subset X$, for all time $t \in [0, T_f]$ and for all $\varepsilon \in [0, \varepsilon^*]$. This general setting applies to many Partial Differential Equations considered as evolution equations in a Banach space. We fully treat below the case of semi-linear wave equations in a high-frequency regime.

In a more technical way, we look for a periodic, near identity and smooth, change of variable $\phi(\frac{t}{\varepsilon}, \cdot)$ and a flow ψ_t of a non-oscillatory autonomous differential equation (1.2) on X such that the solution of the original equation can be written as

$$u(t) = \phi\left(\frac{t}{\varepsilon}, \psi_t(u_0)\right). \quad (1.3)$$

We refer to Lochak-Meunier [39] or Sanders-Verhulst-Murdoch [47] for textbooks on these issues. We point out here, that there are several possible choices for the change of variable ϕ . The choice made in the work of Perko is such that ϕ is the identity map in average. Here, we impose $\phi(0, \cdot) = \text{Id}$ as in [16] and [11]. Therein, it is shown that stroboscopic averaging is the only *geometric* averaging procedure (i.e. structure preserving). It also leads to a numerical method, the so-called Stroboscopic Averaging Method (SAM), introduced in [10].

The formulas given by Perko in [44] and used in section 2.2 can be seen as an expansion in ε of ϕ and G . By differentiating both sides of (1.3) with respect to t , we get

$$\frac{du}{dt} = \frac{1}{\varepsilon} \frac{\partial \phi(t/\varepsilon, \psi_t(u_0))}{\partial t} + D\phi\left(t/\varepsilon, \psi_t(u_0)\right) \left[\frac{\partial \psi_t(u_0)}{\partial t} \right], \quad (1.4)$$

where $D\phi$ denotes the derivatives of ϕ with respect to the variable in X . We formally expand $\phi(\frac{t}{\varepsilon}, V) = V + \varepsilon\phi_1(\frac{t}{\varepsilon}, V) + \varepsilon^2\phi_2(\frac{t}{\varepsilon}, V) + \dots$. Using equations (1.1), (1.2) and relation (1.3), we deduce from (1.4), the expression for G ,

$$G(V; \varepsilon) = \left(D\phi\left(\frac{t}{\varepsilon}, V\right) \right)^{-1} \left[g\left(\frac{t}{\varepsilon}, \phi\left(\frac{t}{\varepsilon}, V\right), \varepsilon\right) - \frac{1}{\varepsilon} \frac{\partial \phi(\frac{t}{\varepsilon}, V)}{\partial t} \right]. \quad (1.5)$$

We then expand $g \circ \phi$ and $D\phi^{-1}$ by their Taylor's formula. For example, $D\phi^{-1}(\frac{t}{\varepsilon}, V) = V - \varepsilon D\phi_1(\frac{t}{\varepsilon}, V) + \mathcal{O}(\varepsilon^2)$ and $g(\frac{t}{\varepsilon}, \phi(\frac{t}{\varepsilon}, V), \varepsilon) = g(\frac{t}{\varepsilon}, V, \varepsilon) + \varepsilon Dg(\frac{t}{\varepsilon}, V, \varepsilon)[\phi_1(t/\varepsilon, V)] +$

$\mathcal{O}(\varepsilon^2)$. By using expression (1.5), the expansions of $g \circ \phi$, $D\phi^{-1}$ and by identifying the terms with identical power in ε we obtain,

$$\begin{aligned} G_1 &= g - \frac{\partial \phi_1}{\partial t} &:= g_1 - \frac{\partial \phi_1}{\partial t}, \\ G_2 &= D\phi_1 g_1 + Dg\phi_1 - \frac{\partial \phi_2}{\partial t} &:= g_2 - \frac{\partial \phi_2}{\partial t}. \end{aligned}$$

This expansion leads to the functions g_n defined in Section 2.2. Then, the change of variables ϕ_n and the vector field G_n are defined respectively by the equation $\frac{\partial \phi_n}{\partial t} = G_n - g_n$ (called homological equation in the Kolmogorov, Arnold, Moser (KAM) context) and the relation $G_n(V, \varepsilon) = \frac{1}{T} \int_0^T g_n(\tau, V, \varepsilon) d\tau$. One key-aspect of our work is to adapt and extend Perko's results to the case of stroboscopic averaging in a Banach space.

Finally, we present and apply the Stroboscopic Averaging Method (SAM) to general non-linear wave equation (NLW) in a high-frequency regime. The SAM was introduced in the ODE case by Calvo, Chartier, Murua and Sanz-Serna in [10] and applied to non-linear Schrödinger equations by Castella, Chartier, Méhats and Murua in [13]. The numerical part of this paper is concerned with two aspects. First, the study of qualitative properties for a resonant NLW (harmonic actions and Fourier modes) and its averaged system. Second, we observe that the SAM displays the expected qualitative properties (conservation of energy, accurate approximation of harmonic actions and Fourier modes, numerical error). While the dynamics of the resonant non-linear Schrödinger equation in the torus and the non-resonant NLW have been studied by various authors (see for instance [3], [6], [23], [12] and [30]), the dynamics of the resonant wave equation has remained mostly unexplored up to our knowledge. This paper is a contribution in this direction.

Our first main theorem (Theorem 2.7), an averaging result in an abstract Banach space generalizing the theorem of Perko (Theorem 2.4), is stated in Subsection 2.5, once preliminary assumptions and Perko's formulas have been settled in Subsection 2.2 and properties of analytic functions in Banach space recalled in Subsection 2.4. Section 3 considers the application for NLW of Theorems 2.4 and 2.7. In the NLW context, Theorem 3.3 (Dirichlet boundary conditions) and Theorem 3.4 (periodic boundary conditions) can be considered as the main output of this paper. Finally, Section 4 is devoted to numerical illustrations obtained by the SAM. The numerical method and a confirmation of the numerical order behaviour are exposed in Subsection 4.2. Subsection 4.3 and 4.4 are devoted to numerical illustrations of the qualitative properties of the SAM.

2 High-order averaging and analytic functions in a Banach space

We present in this section the formulas of high-order stroboscopic averaging and the main theorem of this paper which states the exponentially small error estimates.

2.1 Preliminaries

Let X be a real Banach space equipped with the norm $\|\cdot\|_X$, N an integer and $u_0 \in X$ the initial value. The function g in (1.1) is supposed to satisfy the following:

Assumption 5 (N) *The function $g : \mathbb{T} \times X \times \mathbb{R}_+ \rightarrow X$ is*

- *T -periodic and continuous w.r.t. the first variable,*
- *continuous w.r.t. the third variable,*
- *of class $\mathcal{C}^N$ w.r.t. the second variable and there exists $0 < \varepsilon^* \leq 1$ such that g and all its derivatives are locally bounded in X uniformly in the first and the third variable for $0 \leq \varepsilon \leq \varepsilon^*$.*

Notations For the sake of clarity, we will, unless necessary, omit the third variable in g . We denote the j^{th} derivative of g with respect to u in the directions $a_1, \dots, a_j \in X$ by $D^j g(\tau, u)[a_1, \dots, a_j]$ and for $a \in X$, $a^j := [a, \dots, a]$. Let $\|\cdot\|_{\mathcal{L}_X^j}$ denotes the norm of a continuous j -linear mapping defined as

$$\|D^j g(\tau, u)\|_{\mathcal{L}_X^j} := \sup_{a_1, \dots, a_j \in X} \frac{\|D^j g(\tau, u)[a_1, \dots, a_j]\|_X}{\|a_1\|_X \cdots \|a_j\|_X}.$$

For $j = 0$, by convention, we denote $D^0 g(\tau, u) := g(\tau, u)$ and $\mathcal{L}_X^0 := X$.

Let $r_0 > 0$ and $r \geq 0$. We denote by $B_r \subset X$ the open ball centred in u_0 with radius r . For $j \in \mathbb{N}$, we introduce the norm $\|\cdot\|_r$,

$$\|D^j g\|_r = \sup_{\tau \in \mathbb{T}; u \in B_{r_0+r}, 0 \leq \varepsilon \leq \varepsilon^*} \|D^j g(\tau, u, \varepsilon)\|_{\mathcal{L}_X^j},$$

for the functions g satisfying Assumption 5(j).

In this notations, Taylor's expansion of a function g satisfying Assumption 5(N) with respect to u is written, for $h = u - u_0$, $\tau \in \mathbb{R}$ and $u \in X$,

$$g(\tau, u) = g(\tau, u_0) + \dots + \frac{1}{(N-1)!} D^{N-1} g(\tau, u_0) h^{N-1} + R_N(\tau, u),$$

where

$$R_N(\tau, u) = \frac{1}{(N-1)!} \int_0^1 (1-\theta)^{N-1} D^N g(\tau, u - \theta h) h^N d\theta. \quad (2.1)$$

The following result can be deduced straightforwardly from Grönwall lemma.

Lemma 2.1 *Consider $g : \mathbb{R} \times X \times \mathbb{R}_+ \rightarrow X$ satisfying Assumption 5(1). We suppose that $u_0 \in X$, define the function $G_1 : X \times \mathbb{R}_+ \rightarrow X$ as*

$$G_1(u) = \frac{1}{T} \int_0^T g(\tau, u) d\tau, \quad (2.2)$$

and denote $M_1 := \|DG_1\|_0$. Then, equation

$$\frac{dV_1}{dt}(t) = G_1(V_1(t)), \quad V_1(0) = u_0 \quad (2.3)$$

has a unique solution remaining in B_{r_0} for all, $0 \leq t \leq T_f$ with

$$T_f := \frac{1}{M_1} \ln \left(\frac{r_0}{M_1 \|u_0\|_X} \right).$$

In the sequel, the equation (1.1) will be studied on $[0, T_f]$.

2.2 Perko's formulas

In this subsection, we recall some formulas and results from [44]. As mentioned in the introduction, averaging formulas can be deduced from relation (1.3). The main difference here is that the relations are stated in a Banach space X and in the case of stroboscopic averaging.

We define, for $(t, \varepsilon) \in [0, T_f] \times]0; \varepsilon^*]$, the N th order approximation function u_N by

$$u_N(t) = V_N(t) + \sum_{j=1}^{N-1} \varepsilon^j \phi_j \left(\frac{t}{\varepsilon}, V_N(t) \right), \quad (2.4)$$

where V_N is the solution of the truncated autonomous system

$$\frac{dV_N}{dt} = G_1(V_N) + \dots + \varepsilon^{N-1} G_N(V_N), \quad V_N(0) = u_0, \quad (2.5)$$

and the maps ϕ_j are changes of variables from X to X defined by the formula

$$\phi_j(\tau, u) = \frac{1}{T} \int_0^\tau \left(g_j(\theta, u) - G_j(u) \right) d\theta, \quad (2.6)$$

where, $\tau \in [0, T_f/\varepsilon]$, $u \in X$,

$$G_j(u) = \frac{1}{T} \int_0^T g_j(\tau, u) d\tau, \quad (2.7)$$

and

$$g_j(\tau, u) = \sum_{k=1}^{j-1} \left(\sum_{i_1+\dots+i_k=j-1} \frac{1}{k!} D^k g(\tau, u) [\phi_{i_1}(\tau, u), \dots, \phi_{i_k}(\tau, u)] \right. \\ \left. - D\phi_k(\tau, u) [G_{j-k}(u)] \right). \quad (2.8)$$

Ultimately, we denote $g_1 := g$, the function introduced in equation (1.1). One can remark that equation (2.3) in Lemma 2.1 is the autonomous equation (2.5) with $N = 1$. The difference with [44], is that the changes of variables ϕ_j verify $\phi_j(0, u) = \text{Id}$. In order to prove Theorem 2.4, we now give two lemmas.

Lemma 2.2 *Let $N > 0$ be an integer and suppose that g satisfies Assumption 5 (N). Then, for all $j \in \llbracket 1, N \rrbracket$, the functions ϕ_j , G_j and g_j respectively defined by (2.6), (2.7) and (2.8) verify Assumption 5 ($N - j + 1$).*

Proof of Lemma 2.2 The regularity of ϕ_j , g_j and G_j can be proved easily by an induction argument. The periodicity of ϕ_j with respect to τ stems from

$$\phi_j(\tau + T, u) = \frac{1}{T} \int_T^{\tau+T} (g_j(\theta, u) - G_j(u)) d\theta,$$

and from the periodicity of g_j .

Lemma 2.3 *Let $N > 0$ be an integer, $r \geq 0$ and $0 < \varepsilon \leq \varepsilon^*$. We suppose that g satisfies Assumption 5 (N) and that, equations (1.1) and (2.5) have a unique solution noted respectively $u(t)$ and $V_N(t)$ in B_{r_0+r} for $t \in [0, T_f]$. We define the N^{th} -order approximation*

$$\tilde{u}_N(t) := u_N(t) + \varepsilon^N \phi_N(t/\varepsilon, V_N(t)) \quad (2.9)$$

where u_N and ϕ_N are defined by (2.4) and (2.6). We finally introduce for $(\tau, u) \in \mathbb{T} \times B_{r_0+r}$,

$$\tilde{R}_N(\tau, u) = \sum_{k=1}^{N-1} \sum_{j=N+1}^{kN+1} \sum_{\substack{i_1+\dots+i_k=j-1 \\ i_1, \dots, i_k \leq N}} \frac{\varepsilon^{j-N-1}}{k!} D^k g(\tau, u) [\phi_{i_1}(\tau, u), \dots, \phi_{i_k}(\tau, u)] \quad (2.10)$$

and

$$\bar{R}_N(\tau, u) = \sum_{j=N}^{2N} \varepsilon^{j-N} \sum_{\substack{i_1+i_2=j \\ i_1, i_2 \leq N}} D\phi_{i_1}(\tau, u) G_{i_2}(u). \quad (2.11)$$

Then, we have the following relation for all $t \in [0, T_f]$,

$$\left(\frac{du}{dt} - \frac{d\tilde{u}_N}{dt} \right)(t) = g\left(\frac{t}{\varepsilon}, u(t)\right) - g\left(\frac{t}{\varepsilon}, \tilde{u}_N(t)\right) \\ + \varepsilon^N (R_N + \tilde{R}_N - \bar{R}_N)\left(\frac{t}{\varepsilon}, V_N(t)\right), \quad (2.12)$$

where R_N is defined by expression (2.1). Besides, the functions R_N , $\tilde{R}_N$ and $\bar{R}_N$ are bounded in $\mathbb{T} \times B_{r_0+r}$.

Proof of Lemma 2.3 The proof of relation (2.12) in this lemma uses Taylor expansion and is similar to the proof of Theorem 1 in [44]. Boundedness of R_N , $\tilde{R}_N$ and $\bar{R}_N$ stems directly from relations (2.1), (2.10), (2.11) and Lemma 2.2.

2.3 Averaging theorem with polynomial remainder

Before stating our main result (exponential error estimate), let us prove a polynomial error result.

Theorem 2.4 *Let $N \geq 1$ be an integer, $u_0 \in X$ and $r \geq 0$. We suppose that g satisfies Assumption 5 (N). We denote by $M := \sum_{i=1}^N \|DG_i\|_0$, $\tilde{M} := \sum_{i=2}^N \|G_i\|_0$,*

$$C_1 := (\|R_1\|_r + \|\tilde{R}_1\|_r + \|\bar{R}_1\|_r) e^{M_1 T_f} + \|\phi_1\|_r$$

and

$$\varepsilon_N(r) = \min \left(\varepsilon^*, \frac{r}{2(\tilde{M} T_f) e^{M T_f}}, \frac{r}{2 \sum_{j=1}^{N-1} \|\phi_j\|_{r/2}}, \frac{r}{C_1} \right). \quad (2.13)$$

Then for $0 < \varepsilon \leq \varepsilon_N(r)$, the following statements hold true.

1. Equation (2.5) has a unique solution $V_N(t)$ in $B_{r_0+r/2}$ for $t \in [0, T_f]$.
2. Equation (1.1) has a unique solution $u(t)$ in B_{r_0+r} for $t \in [0, T_f]$.
3. The N^{th} order approximation, $u_N(t)$ defined by (2.4) remains in B_{r_0+r} for $t \in [0, T_f]$ and satisfies,

$$\sup_{t \in [0, T_f]} \|u(t) - u_N(t)\|_X \leq C_N \varepsilon^N, \quad (2.14)$$

where $C_N = (\|R_N\|_r + \|\tilde{R}_N\|_r + \|\bar{R}_N\|_r) e^{M_1 T_f} + \|\phi_N\|_r$ is a constant independent of ε and t .

Proof of Theorem 2.4 By applying Lemma 2.1, equation (2.3) has a unique solution $V_1(t)$ remaining in B_{r_0} for all $0 \leq t \leq T_f$. We now prove the first point of the Theorem.

By standard techniques, it is clear that equation (2.5) has a unique solution for $t \in [0, T_f]$. We now prove that $V_N(t)$ remains in $B_{r_0+r/2}$. Using equations (2.3) and (2.5), we have

$$V_N(t) - V_1(t) = \varepsilon \int_0^t \left(G_1(V_N(\alpha)) + \dots + \varepsilon^{N-1} G_N(V_N(\alpha)) - G_1(V_1(\alpha)) \right) d\alpha,$$

for $t \in [0, T_f]$. This leads to

$$\begin{aligned} \|V_N(t) - V_1(t)\|_X &\leq \varepsilon \int_0^t \sum_{i=1}^N \|G_i(V_N(\alpha)) - G_i(V_1(\alpha))\|_X d\alpha \\ &\quad + \varepsilon \int_0^t \sum_{i=2}^N \|G_i(V_1(\alpha))\|_X d\alpha. \end{aligned} \quad (2.15)$$

By Lemmas 2.1 and 2.2, we have for all $t \in [0, T_f]$ and $i \in \llbracket 1, N \rrbracket$,

$$\|G_i(V_1(t))\|_X \leq \|G_i\|_0 \quad (2.16)$$

and

$$\|G_i(V_N(t)) - G_i(V_1(t))\|_X \leq \|DG_i\|_0 \|V_N(t) - V_1(t)\|_X. \quad (2.17)$$

Using relations (2.13), (2.15), (2.16), (2.17) and Grönwall Lemma, we obtain for $t \in [0, T_f]$,

$$\|V_N(t) - V_1(t)\|_X \leq \varepsilon (\tilde{M} T_f) \exp(M T_f) \leq \frac{r}{2},$$

which proves 1.

For the second and third points, one can remark that by definition of u_N (see (2.4)), we have for all $t \in [0, T_f]$,

$$\|u_N(t) - V_1(t)\|_X \leq \|(V_N - V_1)(t)\|_X + \varepsilon \sum_{j=1}^{N-1} \|\phi_j\|_{r/2}.$$

We conclude that u_N is in B_{r_0+r} using (2.13) for ε_N .

Once again, standard techniques ensure that equation (1.1) has a unique solution $u(t)$ for $t \in [0, T_f]$. We introduce the N^{th} order approximation $\tilde{u}_N$ defined by relation (2.9). We deduce from Lemma 2.3 and relation (2.12) that

$$\left\| \frac{du}{dt}(t) - \frac{d\tilde{u}_N}{dt}(t) \right\|_X \leq M_1 \|u(t) - \tilde{u}_N(t)\|_X + \varepsilon^N [\|R_N\|_r + \|\tilde{R}_N\|_r + \|\bar{R}_N\|_r]$$

for $t \in [0, T_f]$. By Grönwall lemma, we deduce that for $t \in [0, T_f]$

$$\|u(t) - \tilde{u}_N(t)\|_X \leq \varepsilon^N (\|R_N\|_r + \|\tilde{R}_N\|_r + \|\bar{R}_N\|_r) e^{M_1 T_f}.$$

Due to the relation $\|u_N(t) - \tilde{u}_N(t)\|_X \leq \varepsilon^N \|\phi_N\|_r$ we finally get for $t \in [0, T_f]$,

$$\|u(t) - u_N(t)\|_X \leq \varepsilon^N C_N. \quad (2.18)$$

The next and final step is to prove that the solution u remains in B_{r_0+r} for all $t \in [0, T_f]$. For $N = 1$, we have $u_1 = V_1$ and the relation (2.18) holds true. This means that for $t \in [0, T_f]$,

$$\|u(t) - V_1(t)\|_X \leq \varepsilon C_1 \leq r$$

and the result is proved. ■

2.4 Analytic functions in a Banach space

In this part, we deal with analytic functions in a Banach space (see e.g. [45]). We first introduce the complexification of X , defined as

$$X_{\mathbb{C}} = \{U := u + i\tilde{u}, (u, \tilde{u}) \in X^2\}.$$

We denote by $u = \mathcal{R}(U)$ and $\tilde{u} = \mathcal{I}(U)$, the real and imaginary parts of U . The space $X_{\mathbb{C}}$ is a Banach space with the norm

$$\|U\|_{X_{\mathbb{C}}} := \sup_{z \in \mathbb{C}^*} \frac{\|\mathcal{R}(zU)\|_X}{|z|}.$$

Now, given any $\rho > 0$, we consider the (complex) open enlargement of B_{r_0} of size ρ in $X_{\mathbb{C}}$ given by

$$\mathcal{O}_{\rho} = \{u + i\tilde{u} : (u, \tilde{u}) \in B_{r_0} \times X_{\mathbb{C}}, \|\tilde{u}\|_{X_{\mathbb{C}}} < \rho\}.$$

We define analytic functions $f : \mathbb{R} \times X_{\mathbb{C}} \times \mathbb{R}_+ \rightarrow X_{\mathbb{C}}$ on $\mathbb{T} \times \mathcal{O}_{\rho} \times [0, \varepsilon^*]$ as functions continuously differentiable on $\mathcal{O}_{\rho}$ w.r.t. the second variable.

Then, we introduce the norm $\|\cdot\|_{\rho}$ of the bounded analytic functions on $\mathbb{T} \times \mathcal{O}_{\rho}$,

$$\|f\|_{\rho} = \sup_{\tau \in \mathbb{T}, u \in \mathcal{O}_{\rho}, 0 \leq \varepsilon \leq \varepsilon^*} \|f(\tau, u, \varepsilon)\|_{X_{\mathbb{C}}}.$$

Analytic functions on a Banach space share many properties of standard analytic functions of $\mathbb{C}$ like the Cauchy formula. This allows us to use Cauchy's estimates in a neighborhood of any point $u \in \mathcal{O}_{\rho}$, i.e. for any $k \in \mathbb{N}^*$,

$$\|D^k f\|_{\rho-\delta} \leq \frac{k!}{\delta^k} \|f\|_{\rho}. \quad (2.19)$$

We are now ready to formulate the additional assumption for g .

Assumption 6 *The function g is real-analytic in the following sense. There exists $R > 0$ such that g is analytic and bounded by K on $\mathbb{T} \times \mathcal{O}_{2R} \times [0, \varepsilon^*]$.*

The above assumption freezes the definitions of R and K throughout this section. We further introduce $R_j = 2R - j \frac{R}{N}$, for $j \in \llbracket 1, N \rrbracket$.

Lemma 2.5 *If g satisfies Assumption 6, then, for $j \in \llbracket 1, N \rrbracket$, the functions g_j , G_j and ϕ_j , defined by expressions (2.8) (2.7) and (2.6) are analytic and bounded on $\mathbb{T} \times \mathcal{O}_{R_j} \times [0, \varepsilon^*]$ and we have the following inequalities for all k :*

$$\|D^k g\|_{R_j} \leq \frac{k! N^k}{(jR)^k} K, \quad (2.20)$$

$$\|G_j\|_{R_j} \leq \|g_j\|_{R_j}, \quad (2.21)$$

$$\|\phi_j\|_{R_j} \leq 2 \|g_j\|_{R_j}, \quad (2.22)$$

$$\|g_j\|_{R_j} \leq \frac{(Kj)^j N^{j-1}}{R^{j-1}}. \quad (2.23)$$

In order to prove these inequalities, we use the following combinatorial result, whose proof is given in Appendix.

Lemma 2.6 *For two integers $j \geq 1$ and $j \geq k \geq 1$, let*

$$\sigma_k^j = \{(i_1, \dots, i_k) \in \llbracket 1, N \rrbracket^k, \text{ with } i_1 + \dots + i_k = j\}.$$

Then, we have

$$\sum_{(i_1, \dots, i_k) \in \sigma_k^j} i_1^{i_1} \dots i_k^{i_k} \leq j^j \quad (2.24)$$

and

$$\sum_{p=1}^{j-1} p^p (j-p)^{(j-p)} \leq \frac{5}{2} (j-1)^j. \quad (2.25)$$

Proof of Lemma 2.5 Inequality (2.20) is a direct consequence of Cauchy estimates (2.19). Inequalities (2.21) and (2.22) stem directly from (2.6) and (2.7). It remains to prove (2.23). Introducing

$$\alpha_j = \frac{\|g_j\|_{R_j} R^{j-1}}{K^j N^{j-1}}$$

and using the expression (2.8), we get

$$\|g_j\|_{R_j} \leq \sum_{k=1}^{j-1} \left(\sum_{(i_1, \dots, i_k) \in \sigma_k^{j-1}} \frac{1}{k!} \|D^k g\|_{R_j} \|\phi_{i_1}\|_{R_{i_1}} \dots \|\phi_{i_k}\|_{R_{i_k}} \right. \\ \left. + \|D\phi_k\|_{R_k} \|G_{j-k}\|_{R_{j-k}} \right).$$

Then, owing to relations (2.20), (2.21) and (2.22), we obtain

$$\|g_j\|_{R_j} \leq \sum_{k=1}^{j-1} \left(\sum_{(i_1, \dots, i_k) \in \sigma_k^{j-1}} \frac{KN^k}{(jR)^k} (2\|g_{i_1}\|_{R_{i_1}}) \dots (2\|g_{i_k}\|_{R_{i_k}}) \right. \\ \left. + 2\|Dg_k\|_{R_k} \|g_{j-k}\|_{R_{j-k}} \right)$$

and

$$\|g_j\|_{R_j} \leq \frac{K^j N^{j-1}}{R^{j-1}} \sum_{k=1}^{j-1} \left(\sum_{(i_1, \dots, i_k) \in \sigma_k^{j-1}} \left(\frac{2}{j}\right)^k \alpha_{i_1} \dots \alpha_{i_k} + 2\alpha_k \alpha_{j-k} \right).$$

Last inequality then leads to

$$\alpha_j \leq \sum_{k=1}^{j-1} \left(\sum_{(i_1, \dots, i_k) \in \sigma_k^{j-1}} 2^k \alpha_{i_1} \dots \alpha_{i_k} + \frac{2}{j} \alpha_k \alpha_{j-k} \right).$$

Using Lemma 2.6 and an induction argument where we suppose that $\alpha_l \leq l^l$ for all $l < j$, we get

$$\alpha_j \leq \sum_{k=1}^{j-1} \left(\frac{2}{j}\right)^k (j-1)^{j-1} + \frac{2}{j} \times \frac{5}{2} (j-1)^j \\ \leq (j-1)^j \left(\sum_{k=1}^{j-1} \frac{1}{j-1} \left(\frac{2}{j}\right)^k + \frac{5}{j} \right).$$

Notice that for $j \geq 4$, we have

$$\left(\sum_{k=1}^{j-1} \frac{1}{j-1} \left(\frac{2}{j}\right)^k + \frac{5}{j} \right) \leq \frac{1}{j-1} + \frac{5}{j} < 2.$$

The inequalities for $j = 2$ and $j = 3$ are clear. Finally, since $e(j-1)^j \leq j^j$, we get for all j ,

$$\alpha_j \leq j^j$$

which proves (2.23). ■

2.5 Averaging theorem with exponential error estimate

We now present the main theorem of this paper. We use estimates of Theorem 2.4 and Lemma 2.5 to show exponential decrease when the non linearity is analytic.

Theorem 2.7 *Let $u_0 \in X$, $r_0 > 0$, M_1 and T_f be defined as in Lemma 2.1 and suppose that g satisfies Assumption 6. Let $C := (2K + 6R)e^{M_1 T_f}$, $\tilde{C}_1 := \max\left(\frac{K}{R}, \frac{4K}{R^2}\right)$, $\tilde{C}_2 := \frac{\ln 3/2}{\sqrt{3\tilde{C}_1}}$,*

$$\bar{C}_1 := (\|R_1\|_R + \|\tilde{R}_1\|_R + \|\bar{R}_1\|_R)e^{M_1 T_f} + \|\phi_1\|_R,$$

$$\varepsilon_0 := \min\left(\varepsilon^*, \frac{R}{2\bar{C}_1}, \frac{1}{3\tilde{C}_1} \left(\frac{4 - \sqrt{3}}{2}\right)^2, \tilde{C}_2 \ln\left(\frac{2C}{R}\right)\right)$$

and $N(\varepsilon) := \left\lfloor \frac{1}{\sqrt{3\tilde{C}_1\varepsilon}} - 1 \right\rfloor$. Then for all $0 < \varepsilon \leq \varepsilon_0$, the following statements hold true.

1. The solution $u(t)$ of equation (1.1) and the function $u_\varepsilon(t) := u_{N(\varepsilon)}$ defined by (2.4) remain respectively in $\mathcal{O}_{R/2}$ and $\mathcal{O}_R$, for $t \in [0, T_f]$.
2. Function u_ε is exponentially close to function u , i.e.

$$\sup_{t \in [0, T_f]} \|u(t) - u_\varepsilon(t)\|_{X_{\mathbb{C}}} \leq C \exp\left(-\frac{\tilde{C}_2}{\varepsilon}\right).$$

Proof of Theorem 2.7 Following the proof of Theorem 2.4, we have that u satisfies

$$\|u(t) - V_1(t)\|_{X_{\mathbb{C}}} \leq \varepsilon \bar{C}_1 \leq \frac{R}{2},$$

for all $t \in [0, T_f]$. Hence, with standard techniques, there exists $t_1(\varepsilon) > 0$ such that u_ε is well-defined and remains in $\mathcal{O}_R$ for $t \in [0, t_1(\varepsilon)]$.

Here, to simplify the expressions, we denote $N(\varepsilon)$ by N . On the one hand, the definition of $N(\varepsilon)$ and ε_0 give

$$\frac{1}{4\tilde{C}_1\varepsilon} < N^2 < \frac{1}{3\tilde{C}_1\varepsilon}. \quad (2.26)$$

We then deduce from (2.26) that $\varepsilon < 1/(2\tilde{C}_1 N^2)$. On the other hand, relation (2.14) in Theorem 2.4, i.e.

$$\|u(t) - u_N(t)\|_R \leq (\|\phi_N\|_R + (\|\bar{R}_N\|_R + \|\tilde{R}_N\|_R + \|R_N\|_R) \exp(M_1 T_f)) \varepsilon^N,$$

where R_N , $\tilde{R}_N$ and $\bar{R}_N$ are respectively defined by formulas (2.1), (2.10) and (2.11), is valid for all $t \in [0, t_1(\varepsilon)]$. We now give explicit bounds for $\|\phi_N\|_R$, $\|\bar{R}_N\|_R$, $\|\tilde{R}_N\|_R$ and $\|R_N\|_R$. Using Lemma 2.5, we have,

$$\|\phi_N\|_R \leq 2 \frac{(KN)^N N^{N-1}}{R^{N-1}} \leq 2R(\tilde{C}_1 N)^N N^{N-1}.$$

Expressions (2.11), (2.26), Lemma 2.5 and Lemma 2.6 then give,

$$\begin{aligned} \|\bar{R}_N\|_R &\leq \sum_{j=N}^{2N} \varepsilon^{j-N} \sum_{i_1+i_2=j} \|D\phi_{i_1}\|_{R_{i_1}} \|G_{i_2}\|_{R_{i_2}} \\ &\leq \varepsilon^{-N} \sum_{j=N}^{2N} \varepsilon^j 2 \frac{K^j N^{j-1}}{R^{j-1}} \sum_{i_1+i_2=j} i_1^{i_1} i_2^{i_2} \leq 2R\varepsilon^{-N} \sum_{j=N}^{2N} (\tilde{C}_1 \varepsilon j N)^j \\ &\leq \left(\frac{1}{2}\right)^{N-2} R\varepsilon^{-N} \leq 4R(2\tilde{C}_1 N^2)^N. \end{aligned}$$

Similarly, we have,

$$\begin{aligned} \|\tilde{R}_N\|_R &\leq \sum_{k=1}^{N-1} \sum_{j=N+1}^{kN+1} \sum_{i_1+\dots+i_k=j-1} \frac{\varepsilon^{j-N-1}}{k!} \|D^k g\|_{R_1} \|\phi_{i_1}\|_{R_{i_1}} \dots \|\phi_{i_k}\|_{R_{i_k}} \\ &\leq \varepsilon^{-N} \sum_{k=1}^{N-1} \frac{1}{k!} \|D^k g\|_{R_1} \left(\left\| \sum_{i=1}^N \varepsilon^i \phi_i \right\|_{R_i} \right)^k, \end{aligned}$$

and

$$\left\| \sum_{i=1}^N \varepsilon^i \phi_i \right\|_R \leq 2 \sum_{i=1}^N (\varepsilon \tilde{C}_1 i N)^i \leq 2.$$

Moreover, since $\tilde{C}_1 \leq \frac{K}{R}$ and $\frac{R}{4K\varepsilon} \leq N^2$, we obtain

$$\begin{aligned} \|\tilde{R}_N\|_R &\leq \varepsilon^{-N} (2R^{-1})^N K \sum_{k=1}^{N-1} \left(\frac{1}{2}\right)^k \\ &\leq K(2 \times 4KR^{-2}N^2)^N \leq K(2\tilde{C}_1 N^2)^N. \end{aligned}$$

Finally, we compute the last term,

$$\begin{aligned} \|R_N\|_R &\leq \frac{\|D^N g\|_{R_1}}{(N-1)!} \left\| \sum_{j=1}^N \varepsilon^{j-1} \phi_j \right\|_R^N \leq \frac{NK}{R^N} \varepsilon^{-N} \left(\sum_{j=1}^N \|\varepsilon^j \phi_j\|_{R_j} \right)^N \\ &\leq \varepsilon^{-N} (2R^{-1})^N NK \leq K(2\tilde{C}_1)^N N^{2N+1}. \end{aligned}$$

Combining all the terms together, it holds, for all $t \in [0, t_1(\varepsilon)]$,

$$\|u(t) - u_\varepsilon(t)\|_{X_C} \leq C(2\tilde{C}_1 N^2)^{(N+1)}. \quad (2.27)$$

Now, using relation (2.26) and since $N \leq \frac{1}{\sqrt{3\tilde{C}_1\varepsilon}} - 1$, we obtain

$$\|u(t) - u_\varepsilon(t)\|_{X_C} \leq C \left(\frac{2}{3}\right)^{\frac{1}{\sqrt{3\tilde{C}_1\varepsilon}}} \leq C \exp\left(-\frac{\tilde{C}_2}{\varepsilon}\right),$$

for all $t \in [0, t_1(\varepsilon)]$. Finally, since $\varepsilon \leq \tilde{C}_2 \ln\left(\frac{2C}{R}\right)$, the function u_ε remains in B_{r_0+R} for $t \in [0, T_f]$ and stays exponentially close to the function u . $\blacksquare$

2.6 Geometric properties of stroboscopic averaging

As proved in [16] and [13], stroboscopic averaging preserves geometric properties of the original equation. In particular, if g is Hamiltonian, then all averaged vector fields G_j defined by (2.7) are also Hamiltonian. Here, we give a direct proof of this property and exhibit the Hamiltonians corresponding to vector fields G_1 and G_2 , using only formulas (2.6), (2.7) and (2.8).

We assume in this section that X is a Hilbert space equipped with scalar product $(\cdot, \cdot)_X$. Function $g : \mathbb{R} \times X \rightarrow X$ is said to be Hamiltonian (w.r.t. the second variable) if there exists a skew-symmetric, bounded and invertible linear map $J : X \rightarrow X$ and a function $H : \mathbb{T} \times X \rightarrow \mathbb{R}$ (called the Hamiltonian function) such that, for all $\tau \in \mathbb{T}$ and $u \in X$,

$$g(\tau, u) = J^{-1} \nabla H(\tau, u). \quad (2.28)$$

Remark We recall that the gradient in a Hilbert space is defined by

$$\forall (\tau, u, v) \in \mathbb{T} \times X^2, (\nabla H(\tau, u), v)_X := DH(\tau, u)[v].$$

Besides, Lie and Poisson brackets are respectively defined for $g, \tilde{g} : \mathbb{R} \times X \rightarrow X$ and $H, \tilde{H} : \mathbb{R} \times X \rightarrow \mathbb{R}$ by

$$[g(\tau, u), \tilde{g}(\tau, u)] := Dg(\tau, u)[\tilde{g}(\tau, u)] - D\tilde{g}(\tau, u)[g(\tau, u)]$$

and

$$\{H, \tilde{H}\}(\tau, u) := \left(JH(\tau, u), \tilde{H}(\tau, u) \right)_X.$$

It is well known that if g and $\tilde{g}$ are Hamiltonian such that $g(\tau, u) = J^{-1} \nabla H(\tau, u)$ and $\tilde{g}(\tau, u) = J^{-1} \nabla \tilde{H}(\tau, u)$, then the Lie bracket of g and $\tilde{g}$ is also Hamiltonian and satisfies

$$[g(\tau, u), \tilde{g}(\tau, u)] = J^{-1} \nabla \{H, \tilde{H}\}(\tau, u).$$

Before stating the formulas of the Hamiltonian functions corresponding to the averaged flows G_1 and G_2 , we prove the technical lemma:

Lemma 2.8 *Let $g, \tilde{g} : \mathbb{R} \times X \rightarrow X$ two functions satisfying Assumption 5 (1). Then we have*

$$\frac{1}{P} \int_0^P \int_0^\tau \left([g(\tau, u), \tilde{g}(\theta, u)] + [g(\theta, u), \tilde{g}(\tau, u)] \right) d\theta d\tau = F(u)$$

where

$$F(u) = \frac{1}{P} \left[\int_0^P g(\tau, u) d\tau, \int_0^P \tilde{g}(\tau, u) d\tau \right].$$

Proof of Lemma 2.8 With an integration by part, we get directly

$$\begin{aligned} \frac{1}{P} \int_0^P \left[g(\tau, u), \int_0^\tau \tilde{g}(\theta, u) d\theta \right] d\tau &= \frac{1}{P} \left[\int_0^P g(\tau, u) d\tau, \int_0^P \tilde{g}(\tau, u) d\tau \right] \\ &\quad - \frac{1}{P} \int_0^P \left[\int_0^\tau g(\theta, u) d\theta, \tilde{g}(\tau, u) \right] d\tau. \end{aligned}$$

■

Proposition 2.9 *The averaged vector field G_2 defined by (2.7) satisfies*

$$G_2(u) = \frac{1}{P} \int_0^P \frac{1}{2} \left[g(\tau, u), \int_0^\tau g(\theta, u) d\theta \right] d\tau.$$

As a consequence, assuming that the function g satisfies (2.28), the averaged vector fields G_1 and G_2 then are of the form

$$G_1(u) = J^{-1} \nabla H_1(u) \quad \text{and} \quad G_2(u) = J^{-1} \nabla H_2(u),$$

where

$$H_1(u) = \frac{1}{P} \int_0^P H(\tau, u) d\tau,$$

and

$$H_2(u) = \frac{1}{P} \int_0^P \frac{1}{2} \left\{ H(\tau, u), \int_0^\tau H(\theta, u) d\theta \right\} d\tau.$$

Proof of Proposition 2.9 We prove the proposition by using only the formulas of G_2 , g_2 and ϕ_1 defined in Subsection 2.2. We recall that for all $\tau \in \mathbb{T}$ and $u \in X$,

$$\begin{aligned} G_1(u) &= \frac{1}{T} \int_0^T g(\theta, u) d\theta, \\ \phi_1(\tau, u) &= \frac{1}{T} \int_0^\tau (g(\theta, u) - G_1(u)) d\theta \end{aligned} \tag{2.29}$$

and

$$G_2(u) = \frac{1}{T} \int_0^T \left(Dg(\theta, u)\phi_1(\theta, u) - D\phi_1(\theta, u)G_1(u) \right) d\theta.$$

We denote $\dot{\phi}_1(\tau, u) := T \frac{\partial \phi_1}{\partial \tau}(\tau, u) = (g(\tau, u) - G_1(u))$. To simplify, we omit the variables when it is clear. We remark that for all $\tau \in \mathbb{T}$ and $u \in X$,

$$Dg\phi_1 = \frac{1}{2}Dg\phi_1 + \frac{1}{2}DG_1\phi_1 + \frac{1}{2}D\dot{\phi}_1\phi_1, \quad (2.30)$$

and

$$D\phi_1G_1 = \frac{1}{2}D\phi_1G_1 + \frac{1}{2}D\phi_1g - \frac{1}{2}D\phi_1\dot{\phi}_1 \quad (2.31)$$

Using expressions (2.30), (2.31) and the fact that

$$\frac{1}{T} \int_0^T D\dot{\phi}_1(\tau, u)\phi_1(\tau, u)d\tau = -\frac{1}{T} \int_0^T D\phi_1(\tau, u)\dot{\phi}_1(\tau, u)d\tau,$$

we obtain

$$G_2(u) = \frac{1}{T} \int_0^T \frac{1}{2}[g(\tau, u), \phi_1(\tau, u)] + \frac{1}{2}[G_1(u), \phi_1(\tau, u)]d\tau.$$

We have for all $\tau \in \mathbb{T}$ and $u \in X$,

$$[g(\tau, u), \phi_1(\tau, u)] = \left[g(\tau, u), \int_0^\tau g(\theta, u)d\theta \right] - \left[g(\tau, u), \int_0^\tau G_1(u)d\theta \right]$$

and

$$[G_1, \phi_1(\tau, u)] = \left[G_1(u), \int_0^\tau g(\theta, u)d\theta \right].$$

Since, $\frac{1}{T^2} \left[\int_0^T G_1(u)d\tau, \int_0^T g(\tau, u)d\tau \right] = [G_1(u), G_1(u)] = 0$, Lemma 2.8 gives

$$\frac{1}{T} \int_0^T \left[G_1(u), \int_0^\tau g(\theta, u)d\theta \right] - \left[g(\tau, u), \int_0^\tau G_1(u)d\theta \right] d\tau = 0.$$

Then, we deduce

$$G_2(u) = \frac{1}{T} \int_0^T \left[g(\tau, u), \int_0^\tau g(\theta, u)d\theta \right] d\tau$$

and the vector fields G_1 and G_2 derive from the Hamiltonian functions H_1 and H_2 .

■

3 Two examples : Wave equations with Dirichlet and periodic boundary conditions

In this section, we present two examples to illustrate Theorems 2.4 and 2.7. We consider two highly-oscillatory semi-linear wave equations with Dirichlet (section 3.1) and periodic (section 3.2) boundary conditions.

In each case, our goal is to rewrite problem (1.1) under the form

$$\frac{du}{dt} = \frac{1}{\varepsilon}Au + F(u, \varepsilon), \quad u(t=0) = u_0 \in X. \quad (3.1)$$

Here, X will be a Hilbert space and A a skew adjoint unbounded operator from a dense domain $D(A) \subset X$ to X . This ensures, by Stone's theorem (see [43] or [46]) that the operator $\exp(-\tau A)$ is well defined for all $\tau \in \mathbb{R}$. Besides, A will have a compact resolvent and its spectrum will be a subset of $i\lambda\mathbb{N}$ for some $\lambda > 0$. In this case, $\exp(-\tau A)$ is a unitary operator and is periodic with respect to τ .

Then, we rewrite equation (3.1) by applying the change of unknown $v = \exp\left(-\frac{t}{\varepsilon}A\right)u$. We obtain the equation (called the filtered equation)

$$\frac{dv}{dt} = \exp\left(-\frac{t}{\varepsilon}A\right)F\left(\exp\left(\frac{t}{\varepsilon}A\right)v, \varepsilon\right) = g\left(\frac{t}{\varepsilon}, v, \varepsilon\right). \quad (3.2)$$

Finally, we remark that the stroboscopic averaging technique can be applied to other Partial Differential Equations, in particular the ones which can be put under the form (3.1) (for example, the case of a semi linear Schrödinger equation was treated in [13]).

3.1 A semi-linear wave equation with Dirichlet Boundary conditions.

We introduce $\Omega = [0, 1]$ and the highly-oscillatory semi-linear wave equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{1}{\varepsilon^2} \frac{\partial^2 u}{\partial x^2} = \frac{1}{\varepsilon} f(u), \quad t \in \mathbb{R}, \quad x \in \Omega, \quad (3.3)$$

with Dirichlet boundary conditions

$$u(t, x=0) = u(t, x=1) = 0, \quad (3.4)$$

and with initial conditions

$$u(t=0, x) = u_0(x), \quad \varepsilon \frac{\partial u}{\partial t}(t=0, x) = v_0(x). \quad (3.5)$$

Here we have $f : \mathbb{R} \rightarrow \mathbb{R}$ a smooth function satisfying $f(0) = 0$, the initial conditions $u_0 \in H_0^1(\Omega) = \{u \in L^2(\Omega) \text{ s.t. } \partial_x u \in L^2(\Omega) \text{ and } u(0) = u(1) = 0\}$ and $v_0 \in L^2(\Omega)$. We use the norm $\|u\|_{H_0^1} = \|(\sqrt{-\Delta})u\|_{L^2} = \|\partial_x u\|_{L^2}$ on the space H_0^1 . We recall that for any function $u \in L^2(\Omega)$, there exists a unique function $\tilde{u} \in H_0^1(\Omega)$ such that $u = \sqrt{-\Delta}\tilde{u}$. Finally, we denote

$$X = L^2(\Omega) \times L^2(\Omega).$$

Let us put equation (3.3) under the form of equation (3.2). We introduce the new unknown

$$U = \begin{pmatrix} (\sqrt{-\Delta})u \\ \varepsilon \frac{\partial u}{\partial t} \end{pmatrix}.$$

Then U satisfies (3.1) with

$$A = \begin{pmatrix} 0 & \sqrt{-\Delta} \\ -\sqrt{-\Delta} & 0 \end{pmatrix}$$

and

$$F : U = \begin{pmatrix} u \\ v \end{pmatrix} \mapsto \begin{pmatrix} 0 \\ f((\sqrt{-\Delta})^{-1}u) \end{pmatrix}. \quad (3.6)$$

The initial condition associated with this problem, U_0 , belongs to X . Finally, equation (3.3) is equivalent to filtered equation (3.2).

Remark The initial value for the problem (3.2) is now

$$V_0 := U_0 = \begin{pmatrix} (\sqrt{-\Delta})u_0 \\ v_0 \end{pmatrix}.$$

The case with $V_0 = 0$ is trivial. So we assume for the remain of the section that $r_0 := \|V_0\|_X > 0$.

Proposition 3.1 *Assume that f is in $\mathcal{C}^{N+1}(\mathbb{R})$ and satisfies $f(0) = 0$. Then g is well defined and satisfies Assumption 5 (N).*

Proof of Proposition 3.1 Periodicity of g w.r.t. the first variable stems from the properties of the operator A . This is a skew adjoint operator with domain $\mathcal{D}(A) = H_0^1(\Omega) \times H_0^1(\Omega)$ and the embedding $\mathcal{D}(A) \subset X$ is dense. The eigenfunctions and the eigenvalues of $(-\Delta)$ with Dirichlet conditions are respectively $(\sin(\pi kx))_{k \in \mathbb{N}}$ and $(\pi^2 k^2)_{k \in \mathbb{N}}$. Then, the spectrum of A is a subset of $i\pi\mathbb{Z}$, so we can define the operator $\exp(\tau A)$ and remark that this operator is 2-periodic w.r.t. τ and $\|e^{\tau A}\|_{\mathcal{L}_X} \leq 1$, for all $\tau \in \mathbb{R}$.

The well-definition of g and all its N^{th} derivatives come from the following tame estimate. For $\tilde{u} \in H_0^1(\Omega)$ and for $k \in \llbracket 0, N \rrbracket$, the functions $D^k f(\tilde{u})$ are in $\mathcal{L}_{H_0^1(\Omega)}^k$ and

$$\|D^k f(\tilde{u})\|_{\mathcal{L}_{H_0^1}^k} \leq C_f^k(\|\tilde{u}\|_{H_0^1}),$$

where $C_f^k : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ are continuous and non-decreasing functions for all $k \in \llbracket 0, N \rrbracket$.

Now, for $U \in X$, $k \in \llbracket 0, N \rrbracket$ and $\xi_1, \dots, \xi_k \in X$, we denote

$$U := \begin{pmatrix} (\sqrt{-\Delta})\tilde{u} \\ (\sqrt{-\Delta})\tilde{v} \end{pmatrix} \quad \text{and} \quad \forall j \in \llbracket 1, k \rrbracket, \quad \xi_j := \begin{pmatrix} (\sqrt{-\Delta})\tilde{u}_j \\ (\sqrt{-\Delta})\tilde{v}_j \end{pmatrix},$$

where $\tilde{u}, \tilde{v}, \tilde{u}_j$ and $\tilde{v}_j$ are in $H_0^1(\Omega)$. Then F (see (3.6)) satisfies,

$$\begin{aligned} \|D^k F(U)[\xi_1, \dots, \xi_k]\|_X &\leq \|D^k f(\tilde{u})\|_{\mathcal{L}_{H_0^1}^k} \|\tilde{u}_1\|_{H_0^1} \dots \|\tilde{u}_k\|_{H_0^1} \cdot \\ &\leq C_f^k(\|U\|_X) \|\xi_1\|_X \dots \|\xi_k\|_X. \end{aligned}$$

Finally, for all $\tau \in \mathbb{R}$, $V \in X$, $k \in \llbracket 0, N \rrbracket$ and $\xi_1, \dots, \xi_k \in X$,

$$D^k g(\tau, V)[\xi_1, \dots, \xi_k] = e^{-\tau A} D^k F(e^{\tau A} V)[e^{\tau A} \xi_1, \dots, e^{\tau A} \xi_k].$$

Then, $g(\tau, V)$ is in X for all $\tau \in \mathbb{R}$ and $V \in X$, and we have,

$$\|D^k g(\tau, V)[\xi_1, \dots, \xi_k]\|_X \leq \|e^{-\tau A}\|_{\mathcal{L}_X} C_f^k(\|e^{\tau A} V\|_X) \prod_{j=1}^k \|e^{\tau A} \xi_j\|_X,$$

which implies

$$\|D^k g(\tau, V)\|_{\mathcal{L}_X^k} \leq C_f^k(\|V\|_X),$$

since $e^{\tau A}$ is unitary on X . Finally, g satisfies Assumption 5 (N). ■

Proposition 3.2 *Assume that f is real analytic and satisfies $f(0) = 0$. We denote by $\tilde{X} = H_0^1(\Omega) \times H_0^1(\Omega)$, R_0 the radius of analyticity of f and c_1 the norm of the continuous embedding $\tilde{X} \subset X^\infty = L^\infty(\Omega) \times L^\infty(\Omega)$. We finally assume that $r_0 := \|V_0\|_X < \frac{R_0}{c_1}$. Then g satisfies Assumption 6 on $\mathcal{O}_{2R}$ with*

$$\mathcal{O}_{2R} := \{V + \tilde{V} \in B_{r_0} \times X_{\mathbb{C}}, \|\tilde{V}\|_{X_{\mathbb{C}}} < 2R\}$$

and $R > 0$ such that

$$r_0 + 2R < \frac{R_0}{c_1}. \tag{3.7}$$

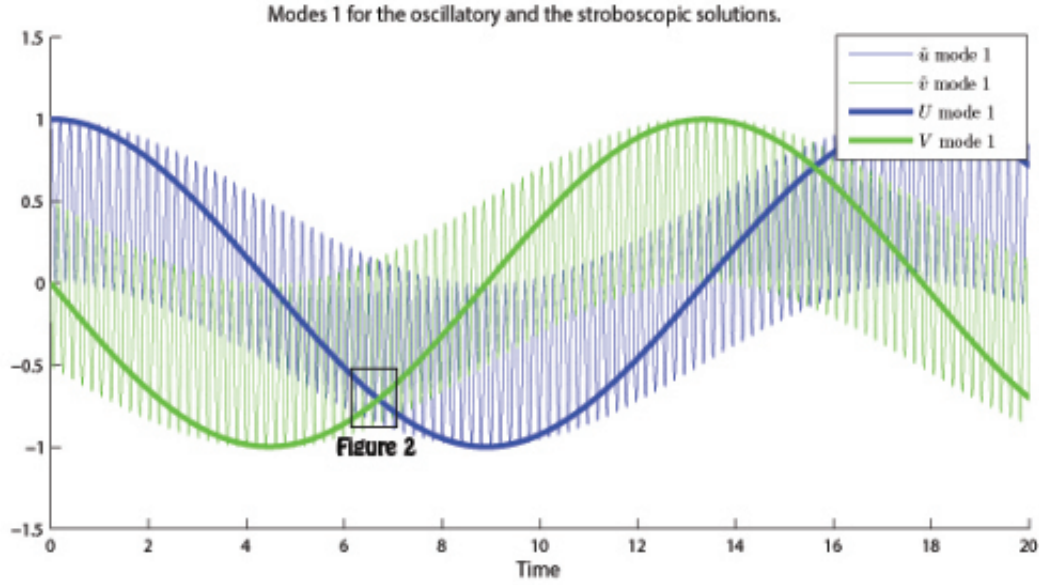


Figure 2.1: The modes 1 for the solutions of equation (3.3) and the corresponding averaged equation (2.5). Computed with $f(u) = -4u^3$, $u_0 = \sin(\pi x)$, $v_0 = 0$ and $\varepsilon = 0.2$.

Proof of Proposition 3.2 For all $U = \begin{pmatrix} (\sqrt{-\Delta})\tilde{u} \\ (\sqrt{-\Delta})\tilde{v} \end{pmatrix}$ in $\mathcal{O}_{2R}$, we have

$$\|\tilde{u}\|_{L^\infty_{\mathbb{C}}} \leq \left\| \begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix} \right\|_{X_{\mathbb{C}}^\infty} \leq c_1(\|U - V_0\|_{X_{\mathbb{C}}} + \|V_0\|_{X_{\mathbb{C}}}) \leq R_0,$$

so $\tilde{u}$ is in the analytic domain of f . Hence we can write, for any function $\tilde{h} \in H_0^1(\Omega)$ such that $\|\tilde{h}\|_{H_{\mathbb{C}}^1}$ (or $\|\tilde{h}\|_{L_{\mathbb{C}}^\infty}$) is small enough, the relation

$$f(\tilde{u} + \tilde{h}) - f(\tilde{u}) - f'(\tilde{u})\tilde{h} = \int_0^1 (1-t)f''(\tilde{u} + t\tilde{h})(\tilde{h})^2 dt.$$

Combined with the fact that $H_0^1(\Omega)$ is an algebra, one can say that the right hand side is an $o(\|\tilde{h}\|_{H_{\mathbb{C}}^1})$. Hence, it is now easy to see that the functions F , then g , are analytic on $\mathcal{O}_{2R}$. Moreover, the boundedness of g on the open bounded set $\mathcal{O}_{2R}$ stems directly from the proof of Proposition 3.1. $\blacksquare$

Theorem 3.3 *Let $N \geq 1$ be an integer, r_0 and R defined as in Proposition 3.2. We assume that $f : \mathbb{R} \rightarrow \mathbb{R}$ is of class $\mathcal{C}^{N+1}$ and $f(0) = 0$, $u_0 \in H_0^1(\Omega)$ and $v_0 \in L^2(\Omega)$. We define $\varepsilon_N(R)$ as in Theorem 2.4. Then for $0 < \varepsilon \leq \varepsilon_N(R)$, the following statements hold true.*

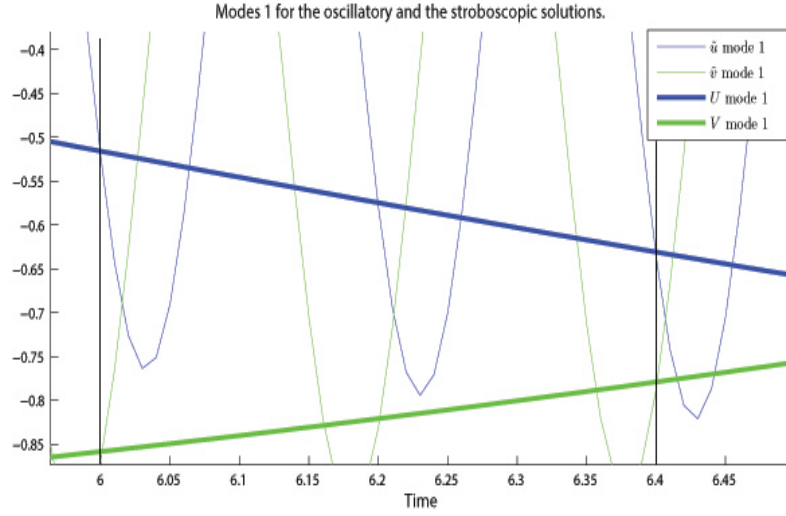


Figure 2.2: Zoom of the figure 2.1 .

1. Equation (3.2) has a unique solution $V(t)$ in B_{r_0+R} for $t \in [0, T_f]$ and equation (3.3) has a unique solution $u(t)$ remaining in a bounded domain for $t \in [0, T_f]$.
2. The N^{th} order averaged equation (2.5) corresponding to equation (3.2) has a unique solution Ψ_N in $B_{r_0+R/2}$ for $t \in [0, T_f]$.
3. The N^{th} order approximation $V_n(t)$ defined by

$$V_N(t) = \Psi_N(t) + \sum_{j=1}^{N-1} \varepsilon^j \phi_j\left(\frac{t}{\varepsilon}, \Psi_N(t)\right),$$

where the changes of variables ϕ_j are defined by (2.6), is in B_{r_0+R} for $t \in [0, T_f]$ and satisfies

$$\sup_{t \in [0, T_f]} \|V(t) - V_N(t)\|_X \leq C_N \varepsilon^N \quad (3.8)$$

with C_N a constant independent of ε and t .

4. Besides, if f is real analytic, we define ε_0 and $N(\varepsilon)$ as in Theorem 2.7. Then, the function $V_\varepsilon := V_{N(\varepsilon)}$ satisfies for $0 < \varepsilon \leq \varepsilon_0$,

$$\sup_{t \in [0, T_f]} \|V(t) - V_\varepsilon(t)\|_{X_{\mathbb{C}}} \leq C \exp\left(-\frac{\tilde{C}_2}{\varepsilon}\right),$$

with $C, \tilde{C}_2$ independent of t and ε .

Proof of Theorem 3.3 Proposition 3.1 ensures that g satisfies Assumption 5 (N). The first 3 statements then follow from Theorem 2.4. Ultimately, if f is real analytic we use Proposition 3.2 and Theorem 2.7 to prove 4. ■

Figure 2.1 represent the first modes (see section 4.2) of the solution of the filtered equation (3.2), together with the solution of the associated averaged equation (2.5). We notice in Figure 2.2 that the exact solution coincide with the stroboscopic solution at stroboscopic times $t_n = 2n\varepsilon$.

3.2 A semi-linear wave equation with periodic boundary conditions.

In this part, we are interested in the semi-linear wave equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{1}{\varepsilon^2} \frac{\partial^2 u}{\partial x^2} = f(u), \quad t \in \mathbb{R}, \quad x \in \Omega, \quad (3.9)$$

with the periodic boundary conditions

$$u(t, x = 0) = u(t, x = 1), \quad (3.10)$$

and the initial conditions

$$u(t = 0, x) = u_0(x), \quad \varepsilon \frac{\partial u}{\partial t}(t = 0, x) = v_0(x). \quad (3.11)$$

Here, the non-linearity f is smooth and $f(0) = 0$, u_0 is in $H_{per}^1(\Omega) := \{u \in L^2(\Omega), \text{ s.t. } \partial_x u \in L^2(\Omega) \text{ and } u(0) = u(1)\}$ and v_0 is in

$$L_0^2(\Omega) := \{u \in L^2(\Omega) \text{ s.t. } \int_0^1 u(y) dy = 0\}.$$

Then, we introduce the Hilbert space

$$X = L_0^2(\Omega) \times L_0^2(\Omega),$$

and the projector $\Pi : H_{per}^1(\Omega) \rightarrow \mathbb{R}$,

$$\Pi : u \rightarrow \int_0^1 u(x) dx.$$

Let us recall that the operator $\sqrt{-\Delta} : H_{per}^1(\Omega) \cap L_0^2(\Omega) \rightarrow L_0^2(\Omega)$ is invertible. So, for $u \in H_{per}^1(\Omega)$, we write $u = u_1 + u_2$ with $u_1 = (\text{Id} - \Pi)u \in H_{per}^1(\Omega) \cap L_0^2(\Omega)$ and $u_2 = \Pi u$.

Equation (3.9) is equivalent to the equations

$$\frac{\partial^2 u_1}{\partial t^2} - \frac{1}{\varepsilon^2} \frac{\partial^2 u_1}{\partial x^2} = f(u_1 + u_2) - \int_0^1 f(u_1 + u_2) dx = f_1(u_1, u_2),$$

$$\frac{\partial^2 u_2}{\partial t^2} = \int_0^1 f(u_1 + u_2) dx = f_2(u_1, u_2), \quad t \in \mathbb{R}, \quad x \in \Omega.$$

Again, we use the change of variables

$$U_1 = \begin{pmatrix} (\sqrt{-\Delta})u_1 \\ \varepsilon \frac{\partial u_1}{\partial t} \end{pmatrix}, \quad U_2 = \begin{pmatrix} u_2 \\ \frac{\partial u_2}{\partial t} \end{pmatrix},$$

to obtain

$$\frac{\partial U_1}{\partial t} = \frac{1}{\varepsilon} A U_1 + \varepsilon F_1(U_1, U_2), \quad t \in \mathbb{R}, x \in \Omega, \quad (3.12)$$

$$\frac{\partial U_2}{\partial t} = F_2(U_1, U_2), \quad (3.13)$$

with, for $\tilde{U}_i = \begin{pmatrix} \tilde{u}_i \\ \tilde{v}_i \end{pmatrix}$ and $i = 1, 2$,

$$F_1 : (\tilde{U}_1, \tilde{U}_2) \rightarrow \begin{pmatrix} 0 \\ f_1((\sqrt{-\Delta})^{-1} \tilde{u}_1, \tilde{u}_2) \end{pmatrix},$$

$$F_2 : (\tilde{U}_1, \tilde{U}_2) \rightarrow \begin{pmatrix} \tilde{v}_2 \\ f_2((\sqrt{-\Delta})^{-1} \tilde{u}_1, \tilde{u}_2) \end{pmatrix},$$

and where A is the skew adjoint operator $\begin{pmatrix} 0 & \sqrt{-\Delta} \\ \sqrt{-\Delta} & 0 \end{pmatrix}$.

With the change of unknown $\tilde{U}_1 = e^{-\frac{t}{\varepsilon} A} U_1$, equation (3.12) is equivalent to

$$\frac{\partial \tilde{U}_1}{\partial t} = \varepsilon \exp\left(-\frac{t}{\varepsilon} A\right) F_1\left(\exp\left(\frac{t}{\varepsilon} A\right) \tilde{U}_1, U_2\right).$$

Finally, equations (3.12) and (3.13) are equivalent to the semi-filtered equation

$$\frac{\partial U}{\partial t} = g\left(\frac{t}{\varepsilon}, U, \varepsilon\right) \quad (3.14)$$

where, for $U = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$,

$$g\left(\frac{t}{\varepsilon}, U, \varepsilon\right) := \begin{pmatrix} \varepsilon \exp\left(-\frac{t}{\varepsilon} A\right) F_1\left(\exp\left(\frac{t}{\varepsilon} A\right) U_1, U_2\right) \\ F_2\left(\exp\left(\frac{t}{\varepsilon} A\right) U_1, U_2\right) \end{pmatrix}.$$

Theorem 3.4 sums up the averaging results for the semi linear wave equation with periodic conditions (3.9).

Theorem 3.4 *Let $u_0 \in H_{per}^1(\Omega)$, $v_0 \in L_0^2(\Omega)$, $\tilde{X} = H_{per}^1(\Omega) \times H_{per}^1(\Omega)$, $U_0 = (\sqrt{-\Delta}u_0, v_0)$, $r_0 := \|U_0\|_X > 0$. Let $R > 0$ and assume that $f : \mathbb{R} \rightarrow \mathbb{R}$ is of class $\mathcal{C}^{N+1}$ and $f(0) = 0$. Then, we define ε_N as in Theorem 2.4 and for $0 < \varepsilon \leq \varepsilon_N$, the followings hold true.*

1. *Equation (3.14) has a unique solution $U(t)$ in B_{r_0+R} for $t \in [0, T_f]$ and Equation (3.9) has a unique solution $u(t)$ remaining in a bounded domain for $t \in [0, T_f]$.*
2. *The N^{th} order averaged equation (2.5) corresponding to (3.14) has a unique solution $\Psi_N(t)$ in $B_{r_0+R/2}$ for $t \in [0, T_f]$.*
3. *The N^{th} order approximation U_n defined by*

$$U_N(t) = \Psi_N(t) + \sum_{j=1}^{N-1} \varepsilon^j \phi_j\left(\frac{t}{\varepsilon}, \Psi_N(t)\right),$$

where the changes of variables ϕ_j are given by (2.6), is in B_{r_0+R} for $t \in [0, T_f]$ and satisfies

$$\sup_{t \in [0, T_f]} \|U(t) - U_N(t)\|_X \leq C_N \varepsilon^N \quad (3.15)$$

with C_N a constant independent of ε and t .

4. *Besides, if f is real analytic, we introduce R_0 its radius of analyticity, c_0 the norm of the continuous embedding $\tilde{X} \subset X^\infty = L^\infty(\Omega) \times L^\infty(\Omega)$, c_1 the norm of the continuous operator $(\sqrt{-\Delta})^{-1} : L_0^2 \rightarrow H_{per}^1 \cap L_0^2$. We assume that ε_0 is defined as in Theorem 2.7 and we have $r_0 > 0$ and $R > 0$ such that,*

$$2c_0c_1r_0 + c_0r_0 < R_0, \quad 2c_0c_1r_0 + c_0r_0 + 2R < R_0.$$

Then g verifies Assumption 6 and the function $U_\varepsilon := U_{N(\varepsilon)}$ satisfies for $0 < \varepsilon \leq \varepsilon_0$,

$$\sup_{t \in [0, T_f]} \|U(t) - U_\varepsilon(t)\|_{X_c} \leq C_1 \exp\left(-\frac{C_2}{\varepsilon}\right)$$

with C_1, C_2 independent of t and ε .

Proof of Theorem 3.4 The proof is similar to the one of Theorem 3.3. We just notice that if f is of class $\mathcal{C}^n$ and $f(0) = 0$, the function g (given by (3.14)) satisfies Assumption 1(N). The difference, here, is the continuous dependence in a third variable ε . For the analytic part of the theorem, we choose M and R such that $(\sqrt{-\Delta})^{-1}u$ is in the analytic domain of f for all $U = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathcal{O}_{2R}$. ■

4 Stroboscopic Averaging Method for a semi-linear wave equation.

We present in this section the Stroboscopic Averaging Method (SAM) introduced by Chartier, Murua and Sanz Serna in [11] and [10]. The SAM is an efficient numerical method for highly-oscillatory ODEs and PDEs. This method is faster than the usual ones to solve problem (1.1) with an acceptable error and preserve some invariants.

4.1 The numerical method

We integrate numerically problem (1.1), with g satisfying Assumption 5(N). We have shown (Theorem 2.4) that the solution V of equation (2.5) and the solution u of problem (1.1) satisfy for all $n \in \mathbb{N}$,

$$u(Tn\varepsilon) = V(Tn\varepsilon) + \mathcal{O}(\varepsilon^N). \quad (4.1)$$

The aim of the SAM is to obtain numerically V at the stroboscopic times $t_n = Tn\varepsilon$, without the (difficult) calculation of the vector field $G^{[N]} := G_1 + \dots + \varepsilon^{N-1}G_N$.

To apply the SAM, we use two numerical methods : One macro-method with a macro-step H and one micro-method with a micro-step h . On the one hand, the micro-method is used to approximate the equation (1.1) over a finite number of period, so the micro-step h has to be smaller than $T\varepsilon$. On the other hand, the macro-step can (and ideally should) be chosen larger than the small period $T\varepsilon$. It is convenient to choose a macro-method that requires only the value $G^{[N]}(u^*; \varepsilon)$ at given values of the argument u^* . In this case, for all $u^* \in X$, we approximate $G^{[N]}(u^*; \varepsilon)$ by a finite difference formula and relation (4.1).

We denote respectively by $\Psi_{t;\varepsilon}$ and $\psi_{t;\varepsilon}$ the flows of equation (1.1) and (2.5). By a finite difference formula, we have

$$G^{[N]}(u^*, \varepsilon) = \left. \frac{\partial}{\partial t} \psi_{t;\varepsilon}(u^*, \varepsilon) \right|_{t=0} = \frac{1}{2\delta} [\psi_{\delta;\varepsilon}(u^*) - \psi_{-\delta;\varepsilon}(u^*)] + \mathcal{O}(\delta^2).$$

Setting $\delta = T\varepsilon$ in the right hand side, we obtain,

$$G^{[N]}(u^*; \varepsilon) = \frac{1}{2T\varepsilon} [\Psi_{2T\varepsilon;\varepsilon}(u^*) - \Psi_{-2T\varepsilon;\varepsilon}(u^*)] + \mathcal{O}(\varepsilon^2) \quad (4.2)$$

To compute $\Psi_{2T\varepsilon;\varepsilon}(u^*)$ and $\Psi_{-2T\varepsilon;\varepsilon}(u^*)$, we integrate problem (1.1) over one period forward and one period backward with the micro-method.

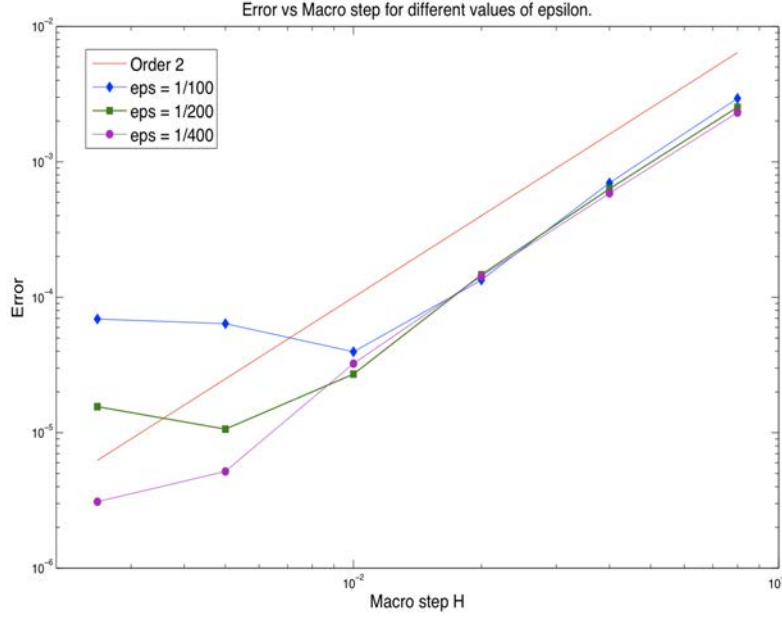


Figure 2.3: Error vs. step size of the macro method for different values of ε .

We can use other difference formulas instead of (4.2). For example, we can approximate $G^{[N]}(u^*, \varepsilon)$ with an error in $\mathcal{O}(\varepsilon^4)$ by using the formula

$$G^{[N]}(u^*; \varepsilon) = \frac{\Psi_{-2T\varepsilon;\varepsilon}(u^*) - 8\Psi_{-T\varepsilon;\varepsilon}(u^*) + 8\Psi_{T\varepsilon;\varepsilon}(u^*) - \Psi_{2T\varepsilon;\varepsilon}(u^*)}{12T\varepsilon} + \mathcal{O}(\varepsilon^4).$$

We make the choice, here, to use symmetric in time approximation of ε because the wave equation is symmetric. Of course, if we have a non symmetric equation or if we can not calculate the solution for reverse times, we use non-symmetric methods.

The error for the SAM is divided in three parts (see [11] or [48]), the approximation of the macro-method, the approximation of G by the finite difference and the approximation of the micro-method. We refer to [11] for a discussion of the various terms of the error. Altogether, one has a bound of the form

$$\mathcal{O}\left(H^p + \varepsilon^\delta + \varepsilon\left(\frac{h}{\varepsilon}\right)^q\right),$$

where p is the order of the macro-method, q the order of the micro-method and δ the order of the finite difference approximation in (4.2). For the numerical tests, we compute the SAM for the semi-linear wave equation (3.3) with Dirichlet boundary conditions and with $f(u) = -4u^3$, $u_0(x) = \sin(\pi x)$ and $v_0(x) = 0$. We use the Runge Kutta method of order 2 for the macro-method and the Strang splitting method for the micro-method. Figure 2.3 illustrates the error versus step size H of the SAM for different values of ε and the results are conform at those expected.

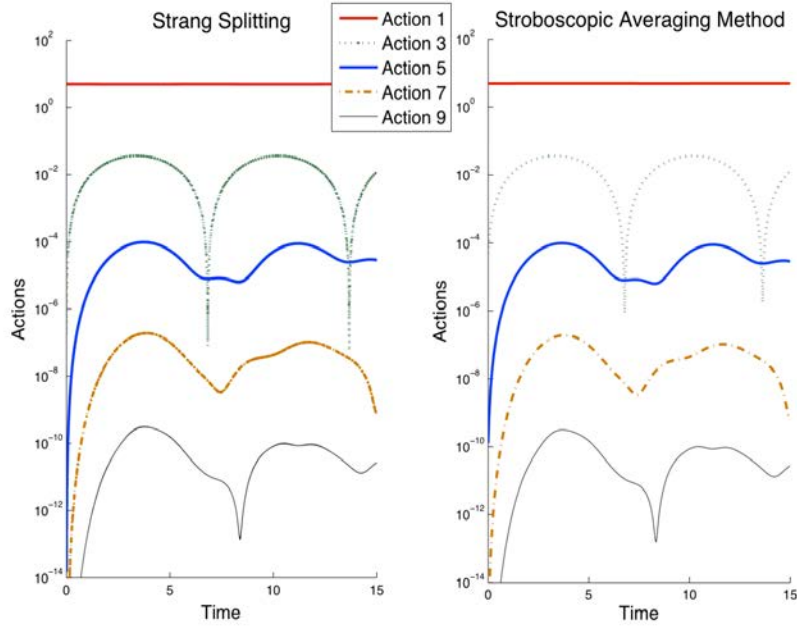


Figure 2.4: Non conservation of the harmonic actions for a resonant semi-linear wave equation.

4.2 A semi-linear wave equation with Dirichlet boundary conditions: Numerical results.

We consider the semi-linear wave equation with Dirichlet boundary conditions (3.3), with $f(u) = -4u^3$ and the initial conditions $u_0(x) = \sin(\pi x)$ and $v_0(x) = 0$. This initial conditions imply that there is only one non-zero harmonic action (see (4.3)) at $t = 0$. Then, at $t > 0$, all the odd harmonic actions are non-zero.

We introduce the frequencies of the system $\omega_n = \pi n$ and the harmonic actions

$$I_n = \frac{1}{2}(\omega_n^2 u_n^2 + v_n^2). \quad (4.3)$$

It is well known that the harmonic actions I_n are constant for the linear wave equation. We also know (see Bambusi [3], Bourgain [6] or Cohen, Hairer, Lubich [23]) that in the case of a non-resonant semi-linear wave equation, the actions remain almost constant for small initial data. The dynamics of the harmonic actions and modes (see below) is not yet investigated for a resonant semi-linear wave equation. We illustrate the behaviour of the actions and modes numerically in figures 2.4, 2.5 and 2.6.

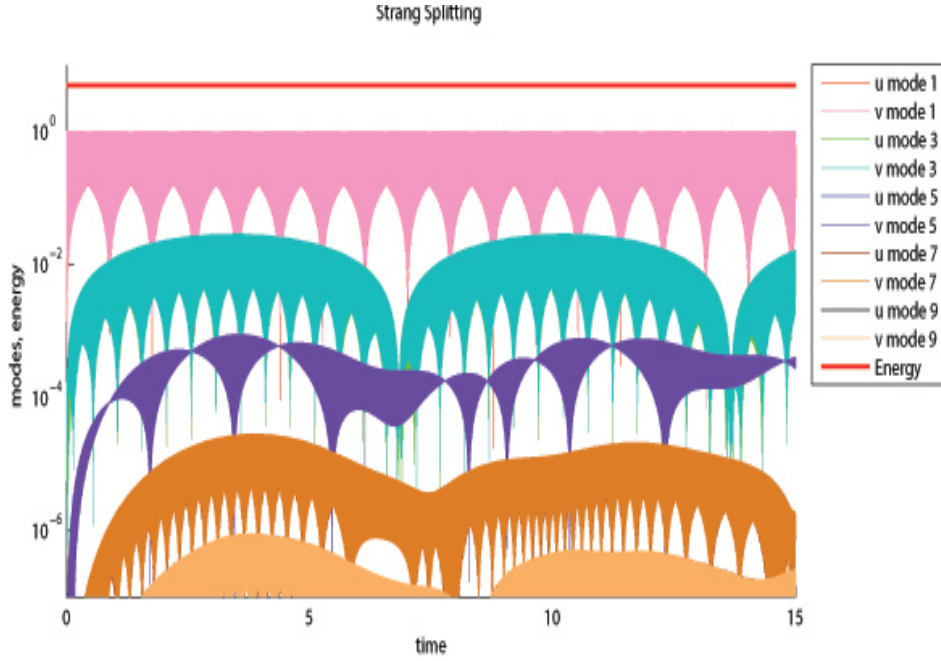


Figure 2.5: Modes for u and v computed with the Strang splitting method, $\varepsilon = 10^{-2}$.

We denote by $u(t)$ the solution of equation (3.3) for $t \in [0, T_f]$ and by v the function $v := \varepsilon \frac{\partial u}{\partial t}$. The functions u and v are in $H_0^1(\Omega)$, so that we have for all $t \in [0, T_f]$, $x \in \Omega$, the Fourier decomposition,

$$u(t, x) = \sum_{n \in \mathbb{N}^*} u_n(t) \sin(\omega_n x)$$

and

$$v(t, x) = \sum_{n \in \mathbb{N}^*} v_n(t) \sin(\omega_n x).$$

The functions u_n and v_n are the n^{th} Fourier modes of the functions u and v . We represent in figure 2.5 different modes for u and v computed with Strang splitting method, $\varepsilon = 10^{-2}$ and an integration step small enough for the numerical solution to be considered as "exact".

Then we compute the modes U_n and V_n for the solution U of the autonomous system (2.5) and the function $V := \varepsilon \frac{dU}{dt}$ in two ways.

- The first order averaged system (2.3) is computed with formula (2.2) where the integral is approximate very accurately (figure 2.6(a)).
- The autonomous system (2.5) is computed with the SAM (figure 2.6(b)).

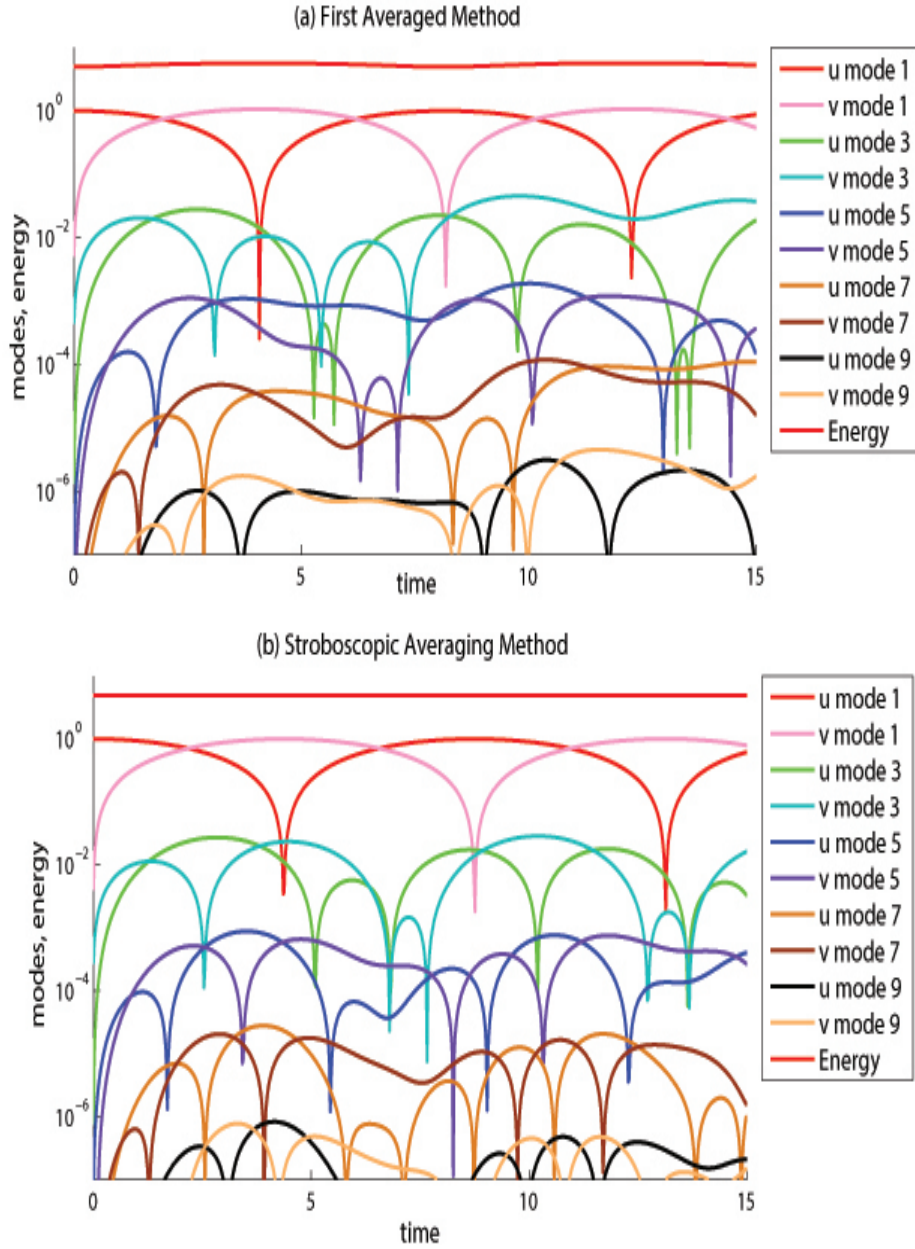


Figure 2.6: The modes of U and V computed with a direct method and with the Stroboscopic Averaging Method.

In regards of these numerical results, we can make the following remarks.

- The dynamics of the modes for the averaged equation is "neater" as compared to the original equation.
- We observe, in Figure 2.6, a beating effect between U_1 and V_1 .
- In contrast with SAM (Figure 2.6(b)), the first order averaged equation (Figure 2.6(a)) does not allow a reliable approximation of modes smaller than ε .

4.3 Energy exchange between modes in a non linear wave equation : heuristic analysis.

In this last part, we look for the formal equations satisfied by the modes u_n , v_n , U_n and V_n . An analysis of the equation for U_1 and V_1 will then help to explain the numerical results and the observed beating effect between U_1 and V_1 .

It is convenient to introduce the functions

$$z_n(t) := \frac{e^{-i\omega_n\tau}}{2} \left(u_n(t) - i \frac{v_n(t)}{\omega_n} \right), \quad (4.4)$$

$z_{-n} := \bar{z}_n$, for all $n \in \mathbb{N}^*$ and we denote by $\tilde{U} := \begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix}$ the solution of equation (3.2), by $\tilde{u}_n$ and $\tilde{v}_n$ the n^{th} modes of $\tilde{u}$ and $\tilde{v}$. Then function z_n can be rewritten as

$$z_n(t) = \frac{1}{2} \left(\tilde{u}_n(t) - i \frac{\tilde{v}_n(t)}{\omega_n} \right).$$

Finally, for all $t \in [0, T_f]$, we introduce

$$Z_n(t) := \frac{1}{2} \left(U_n(t) - i \frac{V_n(t)}{\omega_n} \right) \quad (4.5)$$

and $Z_{-n} := \bar{Z}_n$ for all $n \in \mathbb{N}^*$.

Proposition 4.1 *Let u be the solution of (3.3) with $f(u) = -4u^3$, $u_0(x) = \sin(\pi x)$ and $v_0 = 0$. Then, functions $(z_n)_{n \in \mathbb{Z}^*}$ defined by (4.4) are solutions of the equations*

$$z'_n = -\frac{i}{2\omega_n} \sum_{k,l,m \in \mathbb{Z}^*} z_k z_l z_m e^{i(\omega_k + \omega_l + \omega_m - \omega_n)\pi\tau} \theta_{|k||l||m||n|}, \quad (4.6)$$

with $\theta_{klmn} = \mathbf{1}_{k+l+m=n} + \mathbf{1}_{k-l-m=n} + \mathbf{1}_{l-k-m=n} + \mathbf{1}_{m-l-k=n} - \mathbf{1}_{k-l-m=-n} - \mathbf{1}_{l-k-m=-n} - \mathbf{1}_{m-l-k=-n}$. This implies that the functions $(Z_n)_{n \in \mathbb{Z}^*}$ defined by (4.5) satisfy the equations

$$Z'_n = -\frac{i}{2\omega_n} \sum_{(k,l,m) \in \mathcal{A}_n} Z_k Z_l Z_m \theta_{|k||l||m||n|} + \mathcal{O}(\varepsilon), \quad (4.7)$$

with $\mathcal{A}_n = \{k, l, m \in \mathbb{Z}^*, k + l + m - n = 0\}$.

Proof of proposition 4.1 Replacing the Fourier decomposition of u into equation (3.3), we get,

$$\sum_{n \in \mathbb{N}^*} u_n'' \sin(\omega_n x) + \sum_{n \in \mathbb{N}^*} \left(\frac{\omega_n}{\varepsilon} \right)^2 u_n \sin(\omega_n x) = -\frac{1}{\varepsilon} \left(\sum_{n \in \mathbb{N}^*} u_n \sin(\omega_n x) \right)^3. \quad (4.8)$$

We then develop the right-hand side of (4.8) as

$$-4 \left(\sum_{n \in \mathbb{N}^*} u_n \sin(\omega_n x) \right)^3 = -4 \sum_{k, l, m \in \mathbb{N}^*} u_k u_l u_m \sin(\omega_k x) \sin(\omega_l x) \sin(\omega_m x).$$

This leads to

$$-4 \left(\sum_{n \in \mathbb{N}^*} u_n \sin(\omega_n x) \right)^3 = \sum_{k, l, m, n \in \mathbb{N}^*} u_k u_l u_m \theta_{klmn} \sin(\omega_n x).$$

By projecting on $\sin(\omega_n x)$, we obtain

$$\forall n \in \mathbb{N}^*, \quad u_n'' + \left(\frac{\omega_n}{\varepsilon} \right)^2 u_n = \frac{1}{\varepsilon} \sum_{k, l, m \in \mathbb{N}^*} u_k u_l u_m \theta_{klmn}. \quad (4.9)$$

Now, u_n can be easily written in terms of $\tilde{u}_n$ and $\tilde{v}_n$

$$\begin{aligned} u_n &= \cos(\omega_n \tau) \tilde{u}_n + \frac{1}{\omega_n} \sin(\omega_n \tau) \tilde{v}_n \\ &= \frac{1}{2} \left(\tilde{u}_n - i \frac{\tilde{v}_n}{\omega_n} \right) e^{i\omega_n \tau} + \frac{1}{2} \left(\tilde{u}_n + i \frac{\tilde{v}_n}{\omega_n} \right) e^{-i\omega_n \tau} \\ &= z_n e^{i\omega_n \tau} + z_{-n} e^{-i\omega_n \tau}. \end{aligned}$$

and we get

$$\sum_{k, l, m \in \mathbb{N}^*} u_k u_l u_m \theta_{klmn} = \sum_{k, l, m \in \mathbb{Z}^*} z_k z_l z_m e^{i(\omega_k + \omega_l + \omega_m)\tau} \theta_{|k||l||m|n} := \alpha_n(t, \tau).$$

Equation (4.9) is now equivalent to

$$V_n'(t) = \begin{pmatrix} -\frac{1}{\omega_n} \sin(\omega_n \tau) \alpha_n(t, \tau) \\ \cos(\omega_n \tau) \alpha_n(t, \tau) \end{pmatrix},$$

and for all $n \in \mathbb{N}^*$,

$$z_n' = \frac{1}{2} \left((\tilde{u}_n)' - i \frac{(\tilde{v}_n)'}{\omega_n} \right) = -\frac{i}{2\omega_n} \alpha_n(t, \tau) e^{-i\omega_n \tau} \text{ and } \bar{z}_n' = \frac{i}{2\omega_n} \overline{\alpha_n}(t, \tau) e^{i\omega_n \tau}.$$

Finally, for all $n \in \mathbb{Z}^*$, z_n satisfies equation (4.6). The first order averaged equation then reads as (4.7). $\blacksquare$

A study of the beating effect observed numerically could be interesting. A similar effect was exhibited for NLS (see [30]). However, in our case, this analysis can not be easily adapted. Indeed, all odd modes interact with each other in the first order averaged system (4.7) and it is far from obvious that an expansion in powers of ε can help. Nevertheless, equation (4.7) and the numerical experiments support the following observations:

- One can easily see that the harmonic actions I_n , defined by (4.3), satisfy $I_n = 2\omega_n^2 |Z_n|^2$, with $\omega_n = n\pi$.
- If there are only odd modes in the initial condition, even modes of the solution of the first averaged system and of the oscillatory system (3.3) are zero. In other terms, for $p \in \mathbb{Z}^*$, the solutions z_{2p} of equation (4.6) and Z_{2p} of equation (4.7) vanish.
- When computing the solution U of the averaged equation with SAM, we observe in Figure 2.6(b) that the modes U_{2p+1} and V_{2p+1} remain bounded. The observed bounds for the quantities $\max_{t \in [0, T_f]} (|Z_{2p+1}(t)|)$ for different values of p are gathered in the following table.

p	0	1	2	3	4
$\max(Z_{2p+1})$	0.5	$1.5 \cdot 10^{-2}$	$4.5 \cdot 10^{-4}$	$1.4 \cdot 10^{-5}$	$6 \cdot 10^{-7}$

These numerical results support the following assumption:

$$\exists \delta > 0, \forall p \in \mathbb{N}, \quad |Z_{2p+1}| \leq \frac{1}{2} \delta^p. \quad (4.10)$$

(For this problem, $\delta \approx 0.033$).

For the remaining of the discussion, we then suppose that (4.10) holds true. Since the sum of the harmonic actions remains constant, i.e. $\sum_n I_n(t) = I_1(0) = \frac{\pi^2}{2}$, we have $I_1(t) = \frac{\pi^2}{2} + \mathcal{O}(\delta^2)$. Neglecting all terms bounded by δ in (4.7), we obtain the following equations,

$$\begin{aligned} Z_1' &= \frac{-i}{2\omega_1} (-9|Z_1|^2 Z_1 + \mathcal{O}(\delta)) = iC_1 Z_1 + \mathcal{O}(\delta), \\ Z_3' &= \frac{-i}{2\omega_3} ((Z_1)^3 - 12|Z_1|^2 Z_3 + \mathcal{O}(\delta)), \end{aligned}$$

with $C_1 = \frac{9}{8\pi}$. It follows that

$$Z_1 = \frac{1}{2} e^{iC_1 t} + \mathcal{O}(\delta)$$

and the equation for Z_3 reads

$$Z_3' = iC_2 Z_3 - \frac{i}{48\pi} e^{3iC_1 t} + \mathcal{O}(\delta),$$

for $C_2 = \frac{1}{2\pi}$. We then obtain

$$Z_3(t) = \frac{2i}{48\pi(3C_1 - C_2)} e^{\frac{i(3C_1 + C_2)}{2}t} \sin((3C_1 - C_2)t) + \mathcal{O}(\delta).$$

Now, we remark that $\frac{2}{48\pi(3C_1 - C_2)} \approx 0.015$ is of size δ . As a consequence, function Z_3 is bounded by δ which is consistent with relation (4.10). Let us emphasize that, the solution found for Z_1 in this heuristic analysis is consistent with the numerical tests and Figures 2.1, 2.4 and 2.6.

5 Appendix

Proof of Lemma 2.6 We first prove inequality (2.25) then relation (2.24) is given with an induction argument. We can not prove (2.25) by a simple recurrence, we have to be more precise and remark the symmetry of the expression.

If $j - 1$ is even, then we split the sum like

$$\sum_{p=1}^{j-1} p^p (j-p)^{(j-p)} = 2 \times \sum_{p=1}^{\frac{j-1}{2}} p^p (j-p)^{(j-p)},$$

and we use the fact that $p^p \leq (\frac{j-1}{2})^p$ and $(j-p)^{(j-p)} \leq (j-1)^{(j-p)}$. This gives us

$$\sum_{p=1}^{j-1} p^p (j-p)^{(j-p)} \leq 2 \times (j-1)^j \sum_{p=1}^{\frac{j-1}{2}} \left(\frac{1}{2}\right)^p.$$

The last sum is less or equal to 1 and so we get the result for even $j - 1$.

If $j - 1$ is odd, then the work is the same except that we add a term to the first sum like

$$\sum_{p=1}^{j-1} p^p (j-p)^{(j-p)} = 2 \times \sum_{p=1}^{\frac{j-2}{2}} p^p (j-p)^{(j-p)} + \left(\frac{j}{2}\right)^{\frac{j}{2}} \left(\frac{j}{2}\right)^{\frac{j}{2}}.$$

Then we have,

$$\sum_{p=1}^{j-1} p^p (j-p)^{(j-p)} \leq 2 \times (j-1)^j + \left(\frac{j}{2}\right)^j,$$

and we use the function $x \rightarrow \left(\frac{x}{x-1}\right)^x$ to prove that $\left(\frac{j}{2}\right)^j \leq \frac{1}{2}(j-1)^j$ if $j \geq 3$ and to obtain the result (the case $j = 2$ is obvious).

To prove relation (2.24), we use an induction argument on k and inequality (2.25). If $k = 1$, the result is obvious. Now we remark

$$\sum_{(i_1, \dots, i_{k+1}) \in \sigma_{k+1}^j} i_1^{i_1} \dots i_{k+1}^{i_{k+1}} = \sum_{p=1}^{j-k} p^p \left(\sum_{(i_1, \dots, i_k) \in \sigma_k^{j-p}} i_1^{i_1} \dots i_k^{i_k} \right),$$

and, using the induction property, we get

$$\sum_{(i_1, \dots, i_{k+1}) \in \sigma_{k+1}^j} i_1^{i_1} \dots i_{k+1}^{i_{k+1}} \leq \sum_{p=1}^{j-k} p^p (j-p)^{(j-p)}.$$

Ultimately, we use relation (2.25) to bound the right hand side by $\frac{5}{2}(j-1)^j$. Since $e(j-1)^j \leq j^j$ this completes the proof. ■

Chapter 3

Quasi-stroboscopic averaging in Hilbert spaces

1 Introduction

In this chapter, we are concerned with highly-oscillatory evolution equations posed in a real Hilbert space X (equipped with the scalar product $(\cdot, \cdot)$ and the associated norm $\|\cdot\|$). More precisely, given a fixed $\omega \in \mathbb{R}^d$, we shall consider differential equations in X of the form

$$\frac{d}{dt}u^\varepsilon(t) = \varepsilon g(t, u^\varepsilon(t)), \quad u^\varepsilon(0) = u_0 \in X, \quad t \in [0, T/\varepsilon], \quad (\mathcal{Q}^\varepsilon) \quad (1.1)$$

with $g(t, u) = G(t\omega, u)$ for some function $G : \mathbb{T}^d \times X \rightarrow X$ ($\mathbb{T}^d \equiv (\mathbb{R}/\mathbb{Z})^d$) assumed to be 1-periodic w.r.t to each component of $\mathbb{T}^d$, smooth in $\theta \in \mathbb{T}^d$ and in $u \in X$. Under appropriate assumptions, the solution of (1.1) exists and remains in a bounded path-connected open subset K of X . The main output of this chapter is given in Theorem 2.13. It states that under additional regularity assumptions (analyticity) on the vector field g , the solution u^ε of (1.1) can be obtained, through a change of variables, from the solution of an autonomous *averaged* equation. The specificity of the change of variables retained in the present analysis lies in its stroboscopic nature: it coincides with the identity map at time $t = 0$ and is quasi-periodic. It will be shown in next chapters that this property is of paramount importance for the preservation of geometric structures. In addition, it can be helpful to numerically solve (1.1).

Note that the mono-frequency case has been addressed in previous works, both in finite and infinite dimensional spaces X (see [1] and [13]). The technique developed in [13] in particular can be extended to the new situation at the price of additional assumptions and difficulties. The main new obstacle to which we are confronted here is the occurrence of small denominators in the resolution of the homological equation (for the change of variable). In order to overcome the growth of the coefficients in

Fourier series, we are thus led to require that g is analytic also in the variable t (while continuity is enough in the mono-frequency case). We emphasize that this hypothesis will **not** be satisfied in some situations of interest, such as for instance the Schrödinger equation, where the operator $e^{it\Delta}$ is certainly continuous but sure enough not analytic. However, the technique developed here entails the same advantages as in the mono-frequency case: unlike other papers (Meunier [LM88] or Sanders-Verhulst [SVM07]) where a similar result is derived and where the averaged equation and the change of variables are sought by performing power expansions in ε , we identify here a new equation on the change of variables per se, which can be solved independently using a fixed point procedure. Let us finally emphasize that although not directly applicable, the result derived in this chapter will be used to deal with the Schrödinger equation. The treatment of this equation will combine optimal truncation in ε as explained here and projection onto a space of optimal dimension.

1.1 The formal equations of quasi-stroboscopic averaging

As in the mono-frequency case [13], we first derive formal equations. Generally speaking, the purpose of averaging techniques is to find a near-identity change of variable, quasi-periodic in t (in a sense that will be defined precisely in next section)

$$(t, u) \in \mathbb{R} \times K \mapsto \phi^\varepsilon(t, u) \in X$$

together with the flow map $\psi^\varepsilon(t, u_0)$ of an autonomous differential equation with vector field $\mathcal{G}^\varepsilon$ on X

$$\frac{d}{dt}\psi^\varepsilon(t, u_0) = \varepsilon \mathcal{G}^\varepsilon(\psi^\varepsilon(t, u_0)), \quad (1.2)$$

such that the solution of the original equation (1.1) takes the *composed form*

$$u^\varepsilon(t) = \phi^\varepsilon(t, \psi^\varepsilon(t, (\phi^\varepsilon)^{-1}(0, u_0))). \quad (1.3)$$

In the framework of quasi-stroboscopic averaging, we further impose that the mapping $(t, u) \mapsto \phi^\varepsilon(t, u)$ satisfies $\phi^\varepsilon(0, \cdot) = \text{id}$. This last assumption is similar to the mono-frequency situation. Nonetheless, if $\omega \in \mathbb{R}^d$ is a non-commensurable vector, we emphasize that $\phi^\varepsilon(t, \cdot)$ will not coincide with the identity map for positive time t , a major difference as compared to the mono-frequency case.

Let us now derive (formally) the equations for ϕ^ε and ψ^ε . By differentiating both sides of (1.3) w.r.t. t and using (1.2) we readily get

$$\partial_t \phi^\varepsilon(t, \psi^\varepsilon(t, u_0)) + \varepsilon \partial_u \phi^\varepsilon(t, \psi^\varepsilon(t, u_0)) \mathcal{G}^\varepsilon(\psi^\varepsilon(t, u_0)) = \varepsilon g(t, \phi^\varepsilon(t, \psi^\varepsilon(t, u_0))), \quad (1.4)$$

so that, upon substituting $u_0 = \psi^\varepsilon(-t, u)$ (recall that ψ^ε is the t -flow of an autonomous vector field and as such satisfies $\psi^\varepsilon(-t, \psi^\varepsilon(t, \cdot)) = \text{id}_X$), we obtain

$$\partial_t \phi^\varepsilon(t, u) + \varepsilon \partial_u \phi^\varepsilon(t, u) \mathcal{G}^\varepsilon(u) = \varepsilon g(t, \phi^\varepsilon(t, u)). \quad (1.5)$$

Now, the first equation is obtained by averaging over $t \in \mathbb{R}$ both sides of (1.5). Owing to the quasi-periodicity of ϕ^ε , the average of the term $\partial_t \phi^\varepsilon$ vanishes so that

$$\partial_u \langle \phi^\varepsilon \rangle (u) \mathcal{G}^\varepsilon(u) = \langle g \circ \phi^\varepsilon \rangle (u),$$

where we have used the shortcut notation $g \circ \phi^\varepsilon$ for the function $(t, u) \mapsto g(t, \phi^\varepsilon(t, u))$ and the standard notation

$$\langle f \rangle (u) := \lim_{P \rightarrow +\infty} \frac{1}{P} \int_0^P f(t, u) dt$$

for the average w.r.t. t of a quasi-periodic¹ function $(t, u) \in \mathbb{R} \times X \mapsto f(t, u) \in X$. Assuming for the time being that the linear operator $\partial_u \langle \phi^\varepsilon \rangle (u)$ is invertible for any u , we finally get

$$\mathcal{G}^\varepsilon(u) := (\partial_u \langle \phi^\varepsilon \rangle (u))^{-1} \langle g \circ \phi^\varepsilon \rangle (u). \quad (1.6)$$

In other terms, we have here obtained the value of the vector field $\mathcal{G}^\varepsilon$ (hence that of ψ^ε) in terms of ϕ^ε .

The second equation is then derived by inserting previous relation in equation (1.5)

$$\partial_t \phi^\varepsilon(t, u) + \varepsilon \partial_u \phi^\varepsilon(t, u) (\partial_u \langle \phi^\varepsilon \rangle (u))^{-1} \langle g \circ \phi^\varepsilon \rangle (u) = \varepsilon g(t, \phi^\varepsilon(t, u)). \quad (1.7)$$

This is a closed nonlinear equation on ϕ^ε . However, there is here a sharp discrepancy with the mono-frequency case, as we solve it in the set of quasi-periodic functions and thus have to resort to Fourier expansions. This draws additional technical difficulties and is the very reason why we shall consider functions that are real-analytic in both t and u , and quasi-periodic in t . It thus remains to turn the above formal computations into a rigorous analytic procedure. To this aim, we right-away define the following operators $\Lambda(\phi)$ and $\Gamma^\varepsilon(\phi)$ (whenever $\partial_u \phi$ is invertible):

$$\Lambda(\phi)(t, u) := g(t, \phi(t, u)) - \partial_u \phi(t, u) \langle \partial_u \phi \rangle^{-1}(u) \langle g \circ \phi \rangle (u) \quad (1.8)$$

and

$$\Gamma^\varepsilon(\phi)(t, u) := u + \varepsilon \int_0^t \Lambda(\phi)(\tau, u) d\tau. \quad (1.9)$$

We then can write the so-called homologic equation

$$\partial_t \phi(t, u) = \varepsilon \Lambda(\phi)(t, u), \quad (1.10)$$

or in an integral version

$$\phi(t, u) = \Gamma^\varepsilon(\phi)(t, u). \quad (1.11)$$

¹It is known that for quasi-periodic functions, this limit exists and is finite.

2 Abstract settings

2.1 Analytic functions in a Hilbert space

Given the *real* Hilbert space X , we introduce its complexification defined as

$$X_{\mathbb{C}} = \{U := u + i\tilde{u}, (u, \tilde{u}) \in X^2\},$$

for which we denote $u = \Re(U) \in X$ and $\tilde{u} = \Im(U) \in X$ the real and imaginary parts of U . The space $X_{\mathbb{C}}$ is again a Hilbert space when endowed with the Hermitian norm associated to the complex scalar product

$$(u + i\tilde{u}, v + i\tilde{v})_{X_{\mathbb{C}}} = (u, v) + (\tilde{u}, \tilde{v}) + i((\tilde{u}, v) - (u, \tilde{v})),$$

i.e. the norm

$$\|U\|_{X_{\mathbb{C}}} = \sqrt{\|\Re(U)\|^2 + \|\Im(U)\|^2}.$$

Since for $U \in X$, $\|U\|_{X_{\mathbb{C}}} = \|U\|$, we shall also denote by $\|\cdot\|$ the norm on $X_{\mathbb{C}}$. Now, given any $\rho > 0$, we consider the open enlargement of K in $X_{\mathbb{C}}$ given by

$$K_{\rho} = \{u + \tilde{u} : (u, \tilde{u}) \in K \times X_{\mathbb{C}}, \|\tilde{u}\| < \rho\}.$$

Note that since K is open path-connected in X , so is K_{ρ} in $X_{\mathbb{C}}$. We then define analytic functions as follows:

Definition 2.1 (Analytic functions on X) *Consider a continuous function $u \in K_{\rho} \mapsto f(u) \in X_{\mathbb{C}}$. It is said to be analytic on $K_{\rho} \subset X_{\mathbb{C}}$ if it is continuously differentiable on K_{ρ} , i.e. if there exists a continuous map*

$$\begin{aligned} K_{\rho} &\rightarrow \mathcal{L}(X_{\mathbb{C}}) \\ u &\mapsto (\partial_u f)(u) \end{aligned}$$

where $\mathcal{L}(X_{\mathbb{C}})$ is the set of bounded linear maps from $X_{\mathbb{C}}$ to $X_{\mathbb{C}}$, which satisfies

$$\forall u \in K_{\rho}, \exists \delta > 0, \forall h \in X_{\mathbb{C}}, \|h\| \leq \delta, \|f(u+h) - f(u) - (\partial_u f)(u)h\| = o(\|h\|).$$

When $u \mapsto f(u)$ is a bounded analytic function on K_{ρ} , we further denote

$$\|f\|_{\rho} = \sup_{u \in K_{\rho}} \|f(u)\|.$$

Proposition 2.2 *The set of bounded analytic functions on K_{ρ} , denoted $(\mathbf{A}_{\rho}, \|\cdot\|_{\rho})$, is a Banach space.*

Proof. Consider a Cauchy sequence $(f_n)_{n \in \mathbb{N}}$ in $\mathbf{A}_\rho^\mathbb{N}$ for the norm $\|\cdot\|_\rho$. The space $X_\mathbb{C}$ being complete, it converges pointwise towards a function f . The convergence is furthermore uniform: let $\eta > 0$ be given, $N = N(\eta)$ such that $\|f_n - f_m\|_\rho \leq \eta$ as soon as $m \geq n \geq N$, and consider for any $u \in K_\rho$, $m = m(u) \geq N$ such that $\|f(u) - f_m(u)\| \leq \eta$, then

$$\|f(u) - f_n(u)\| \leq \|f(u) - f_m(u)\| + \|f_m(u) - f_n(u)\| \leq 2\eta$$

so that $\|f - f_n\|_\rho \leq 2\eta$. This implies that f is continuous and bounded on K_ρ and by Morera's theorem, analytic on K_ρ (which is path-connected in $X_\mathbb{C}$). As a matter of fact, each f_n of the Cauchy sequence being holomorphic, for all linear forms $l : X_\mathbb{C} \rightarrow \mathbb{C}$ and all $(u_1, u_2) \in K_\rho \times K_\rho$, the function $z \in \mathbb{C} \mapsto l \circ f_n(u_1 + zu_2) \in \mathbb{C}$ is holomorphic (*a concept known as weak analyticity*) in the usual sense. This implies that for all C^1 closed curve γ such that $\{u_1 + zu_2, z \in \gamma\} \subset K_\rho$

$$\int_\gamma l \circ f_n(u_1 + zu_2) dz = 0$$

and by uniform convergence

$$\int_\gamma l \circ f(u_1 + zu_2) dz = \int_\gamma \left(\lim_{n \rightarrow +\infty} l \circ f_n \right) (u_1 + zu_2) dz = \lim_{n \rightarrow +\infty} \int_\gamma l \circ f_n(u_1 + zu_2) dz = 0.$$

By Morera's Theorem, this proves that $u \mapsto l \circ f(u)$ is analytic on K_ρ and thus f is analytic as well. $\blacksquare$

2.2 Quasi-periodic functions with values in $\mathbf{A}_\rho$

According to Moser [42], quasi-periodic functions $f \in \mathcal{C}^{d+1}(\mathbb{R}, \mathbf{A}_\rho)$ are functions which can be written as²

$$f(t) = F(t\tilde{\omega})$$

for a given vector of frequencies $\tilde{\omega} = (\tilde{\omega}_1, \dots, \tilde{\omega}_d) \in \mathbb{R}^d$ and a given function $F \in \mathcal{C}^{d+1}(\mathbb{T}^d, \mathbf{A}_\rho)$. This vector being specified in the original equation, we shall consider functions f in the set $\mathcal{A}_\rho^\omega$ defined as follows.

Definition 2.3 *A function $f \in \mathcal{C}^{d+1}(\mathbb{R}, \mathbf{A}_\rho)$ belongs to the set $\mathcal{A}_\rho^\omega$ if there exists $F \in \mathcal{C}^{d+1}(\mathbb{T}^d, \mathbf{A}_\rho)$ such that*

$$\forall t \in \mathbb{R}, \quad f(t) = F(t\omega).$$

Functions of $\mathcal{A}_\rho^\omega$ can be expanded in Fourier series. This follows from the following lemma, that we present for the sake of clarity. Prior to that, we define $|x| = \sum_{j=1}^d |x_j|$ for $x \in \mathbb{C}^d$, a notation which will be used in the sequel both for $\theta \in \mathbb{C}^d$ and $k \in \mathbb{Z}^d$.

²Whenever convenient, we shall also write $F(t\tilde{\omega}_1, \dots, t\tilde{\omega}_d)$ in lieu of $F(t\tilde{\omega})$.

Lemma 2.4 *Let f be a function of $\mathcal{A}_\rho^\omega$. Then there exists $F \in \mathcal{C}^{d+1}(\mathbb{T}^d, \mathbf{A}_\rho)$ such that*

$$f(t) = \sum_{k \in \mathbb{Z}^d} e^{2i\pi t(\omega \cdot k)} \hat{F}_k$$

where $\hat{F}_k = \int_{\mathbb{T}^d} e^{-2i\pi(\theta \cdot k)} F(\theta) d\theta$.

Proof. For $f \in \mathcal{A}_\rho^\omega$, there exists $F \in \mathcal{C}^{d+1}(\mathbb{T}^d, \mathbf{A}_\rho)$ such that $f(t) = F(t\omega)$. By assumption, $\theta \in \mathbb{T}^d \mapsto F(\theta) \in \mathbf{A}_\rho$ is $(d+1)$ -times continuously differentiable, hence we may estimate Fourier coefficients (k is here a non-zero multi-index of $\mathbb{Z}^d$)

$$\hat{F}_k = \int_{\mathbb{T}^d} e^{-2i\pi(\theta \cdot k)} F(\theta) d\theta$$

through several integration by parts as follows

$$\|F_k\|_\rho = \left\| \int_{\mathbb{T}^d} \frac{e^{-2i\pi(\theta \cdot k)}}{(-2i\pi k_j)^{d+1}} \partial_{\theta_j}^{d+1} F(\theta) d\theta \right\|_\rho \leq \frac{d^{d+1}}{(2\pi|k|)^{d+1}} \sup_{|\alpha|=d+1, \theta \in \mathbb{T}^d} \|\partial_{\theta_\alpha}^{|\alpha|} F\|_\rho \quad (2.1)$$

where j is an index such that $|k_j| = \sup_{l=1, \dots, d} |k_l|$. The uniform convergence of the series $\sum_{k \in \mathbb{Z}^d} e^{2i\pi(\theta \cdot k)} \hat{F}_k$ towards $F(\theta)$ is then ensured by the summability of $\sum_{k \in \mathbb{Z}^d} |k|^{-d-1}$. $\blacksquare$

2.3 Analytic quasi-periodic functions with values in $\mathbf{A}_\rho$

As already noticed in Introduction, our analysis requires analyticity assumptions with respect to the variable $\theta \in \mathbb{T}^d$. To this aim, given $\sigma > 0$, we introduce the open complex strip D_σ as the enlargement of the d -dimensional torus $\mathbb{T}^d$ defined by

$$D_\sigma = \left\{ \theta = \theta_1 + i\theta_2, (\theta_1, \theta_2) \in \mathbb{T}^d \times \mathbb{R}^d, \|\theta_2\|_\infty < \sigma \right\}. \quad (2.2)$$

Definition 2.5 (Analytic functions with values in $\mathbf{A}_\rho$) *Consider a function $f \in \mathcal{A}_\rho^\omega$. It is said to be analytic on D_σ if the associated function $F \in \mathcal{C}^{d+1}(\mathbb{T}^d, \mathbf{A}_\rho)$ is continuously differentiable on D_σ , i.e. if there exists a continuous map*

$$\begin{aligned} D_\sigma &\rightarrow \mathcal{L}(D_\sigma, \mathbf{A}_\rho) \\ \theta &\mapsto (\partial_\theta F)(\theta) \end{aligned}$$

where $\mathcal{L}(D_\sigma, \mathbf{A}_\rho)$ is the set of bounded linear maps from D_σ to $\mathbf{A}_\rho$, which satisfies

$$\forall \theta \in D_\sigma, \exists \delta > 0, \forall h \in D_\sigma,$$

$$\|h\|_\infty \leq \delta, \sup_{\theta \in \mathbb{T}^d} \|F(\theta + h) - F(\theta) - (\partial_\theta F)(\theta)h\|_\rho = o(\|h\|_\infty).$$

We then denote by $\mathcal{A}_{\delta,\rho}^\omega$ the set of functions $f \in \mathcal{A}_\rho^\omega$ which are analytic on D_ρ for which F is bounded and we define the norm on $\mathcal{A}_{\delta,\rho}^\omega$ by

$$\|f\|_{\sigma,\rho} = \sup_{\theta \in D_\sigma} \|F(\theta)\|_\rho. \quad (2.3)$$

The set $(\mathcal{A}_{\delta,\rho}^\omega, \|\cdot\|_{\sigma,\rho})$ is then a Banach space (a consequence of Morera's theorem). Note that for functions of $\mathcal{A}_{\delta,\rho}^\omega$, the estimate on Fourier coefficients (2.1) of Lemma 2.4 can be refined as

$$\|\hat{F}_k\|_\rho \leq \|f\|_{\sigma,\rho} e^{-2\pi\sigma|k|} \quad (2.4)$$

through a contour-integral in D_σ .

Remark 2.6 Denoting $\tilde{\theta} = (\theta_1, \dots, \theta_{d-1})$, $\tilde{k} = (k_1, \dots, k_{d-1})$ and $0 < \tilde{\sigma} < \sigma$, we have indeed

$$\hat{F}_k = \int_{\mathbb{T}^{d-1}} e^{-2i\pi(\tilde{k} \cdot \tilde{\theta})} \left(\int_0^1 e^{-2i\pi k_d \theta_d} F(\tilde{\theta}, \theta_d) d\theta_d \right) d\tilde{\theta}$$

and the inner-most integral can be cut into three pieces (with $s_d = \pm 1$ and $s = (s_1, \dots, s_d)$)

$$\int_0^{s_d \tilde{\sigma}} i e^{2\pi k_d y} F(\tilde{\theta}, iy) dy + \int_0^1 e^{-2i\pi k_d (x + s_d i \tilde{\sigma})} F(\tilde{\theta}, x + s_d i \tilde{\sigma}) dx + \int_{s_d \tilde{\sigma}}^0 i e^{2\pi k_d y} F(\tilde{\theta}, 1 + iy) dy.$$

By periodicity of F in its last variable θ_d , the first and third integrals cancel each other leaving only the second one. Now repeating this operation d times for each component of θ , we obtain

$$\hat{F}_k = e^{2\pi \tilde{\sigma}(s \cdot k)} \int_{\mathbb{T}^d} e^{-2i\pi(k \cdot \theta)} F(\theta + i\tilde{\sigma}s) d\theta$$

which can be bounded by $e^{-2\pi|k|\tilde{\sigma}} \|f\|_{\sigma,\rho}$ by choosing s_j of the sign of $-k_j$ for all $i = j, \dots, d$. We then get (2.4) in the limit $\tilde{\sigma} \rightarrow \sigma$.

In the process of averaging the following elements of calculus will be frequently used for the average of a function $f \in \mathcal{A}_{\delta,\rho}^\omega$:

$$\langle f \rangle = \lim_{P \rightarrow +\infty} \frac{1}{P} \int_0^P f(t) dt.$$

Lemma 2.7 If f belongs to $\mathcal{A}_{\delta,\rho}^\omega$, then for any $0 < \tilde{\sigma} < \sigma$, $\partial_t f$ belongs to $\mathcal{A}_{\sigma-\tilde{\sigma},\rho}^\omega$ and one has

$$\langle \partial_t f \rangle = 0.$$

Proof. For f in $\mathcal{A}_{\delta,\rho}^\omega$, there exists $F \in \mathcal{C}^{d+1}(\mathbb{T}^d, \mathbf{A}_\rho)$ analytic and bounded on D_σ , such that $f(t) = F(t\omega)$. It follows that $(\partial_t f)(t) = (\omega \cdot \partial_\theta F)(t\omega)$. Since F is analytic on D_σ , each partial derivative $\partial_{\theta_j} F$ is also analytic and bounded on $D_{\sigma-\tilde{\sigma}}$ by $\frac{1}{\tilde{\sigma}} \sup_{\theta \in D_\sigma} \|F(\theta)\|_\rho$ owing to Cauchy's formula. It then follows that

$$\|\partial_t f\|_{\sigma-\tilde{\sigma},\rho} \leq \frac{|\omega|}{\tilde{\sigma}} \|f\|_{\sigma,\rho}$$

where we recall that $|\omega| = \sum_{j=1}^d |\omega_j|$. This implies the first part of the statement. Now,

$$\langle \partial_t f \rangle = \lim_{P \rightarrow +\infty} \frac{1}{P} \int_0^P (\partial_t f)(t) dt = \lim_{P \rightarrow +\infty} \frac{1}{P} (f(P) - f(0)) = 0$$

since $t \mapsto f(t)$ is bounded. ■

Next lemma requires an additional assumption on ω , the so-called ‘‘Siegel’s condition’’ or ‘‘strong non-resonance condition’’ (see [42], [44] or [32]). It is essential in order to deal with small divisors.

Assumption 2.8 *There exist constants $\gamma > 0$, $\nu > 0$, such that*

$$|k \cdot \omega| \geq \gamma |k|^{-\nu},$$

for all k in $\mathbb{Z}^d$ and $k \neq 0$.

Remark 2.9 *It is known (see [42]) that for $\nu > d - 1$, the set of vectors ω which do not satisfy Assumption 2.8 has Lebesgue measure 0 in $\mathbb{R}^d$.*

Lemma 2.10 *Suppose that ω satisfies Assumption 2.8. If f belongs to $\mathcal{A}_{\delta,\rho}^\omega$, then for any $0 < \tilde{\sigma} < \sigma$ the function defined by*

$$\tilde{f}(t) = \int_0^t f(s) ds - t \langle f \rangle$$

belongs to $\mathcal{A}_{\sigma-\tilde{\sigma},\rho}^\omega$. Furthermore, the following estimate holds true

$$\|\tilde{f}\|_{\sigma-\tilde{\sigma},\rho} \leq \frac{\kappa_0}{\tilde{\sigma}^{\nu+d}} \|f\|_{\sigma,\rho} \tag{2.5}$$

with $\kappa_0 = \frac{\nu! 4^d}{\gamma \pi^{\nu+d+1}}$.

Proof. By assumption, $f(t) = \sum_{k \in \mathbb{Z}^d} e^{2i\pi t(k \cdot \omega)} \hat{F}_k$ where the coefficients $\hat{F}_k$ satisfy (2.4). Hence, we have

$$\tilde{f}(t) = \sum_{k \neq 0} \frac{e^{2i\pi t(k \cdot \omega)} - 1}{2i\pi(k \cdot \omega)} \hat{F}_k$$

where the convergence of the series for $t \in \mathbb{R}$ is ensured by

$$\sup_{\theta \in D_{\sigma-\tilde{\sigma}}} \left\| \sum_{k \neq 0} \frac{e^{2i\pi(k \cdot \theta)} - 1}{2i\pi(k \cdot \omega)} \hat{F}_k \right\|_{\rho} \leq \sum_{k \neq 0} \frac{e^{2\pi(\sigma-\tilde{\sigma})|k|} + 1}{2\pi|k \cdot \omega|} \|\hat{F}_k\|_{\rho} \leq \sum_{k \neq 0} \frac{e^{-2\pi\tilde{\sigma}|k|}}{\pi|k \cdot \omega|},$$

which, owing to Assumption 2.8, becomes

$$\|\tilde{f}\|_{\sigma-\tilde{\sigma}, \rho} \leq \frac{\|f\|_{\sigma, \rho}}{\pi\gamma} \sum_{k \neq 0} |k|^{\nu} e^{-2\pi|k|\tilde{\sigma}} \leq \frac{\nu!}{\pi\gamma(\pi\tilde{\sigma})^{\nu}} \sum_{k \neq 0} e^{-\pi|k|\tilde{\sigma}}$$

where we have used that $x^{\nu}/\nu! \leq e^x$ with $x = \pi\tilde{\sigma}|k|$. Finally, the bound (2.5) follows from

$$\sum_{k \in \mathbb{Z}^d} e^{-\pi|k|\tilde{\sigma}} \leq \left(1 + 2 \sum_{j=1}^{+\infty} e^{-j\pi\tilde{\sigma}} \right)^d \leq \left(\frac{1 + e^{-\pi\tilde{\sigma}}}{1 - e^{-\pi\tilde{\sigma}}} \right)^d \leq \left(\frac{4}{\pi\tilde{\sigma}} \right)^d.$$

■

2.4 Main Result: Exponential estimates for the quasi-stroboscopic averaging procedure

We are now in position to state the assumptions on g in (1.1) required by our analysis.

Assumption 2.11 *The mapping g belongs to $\mathcal{A}_{2W, 2R}^{\omega}$ and is bounded by C , i.e.*

$$\|g\|_{2W, 2R} \leq C,$$

where R , W and C are positive constants.

Under Assumption 2.11, it is straightforward to prove the following theorem:

Theorem 2.12 *The Cauchy problem (1.1) is uniformly well-posed in the following sense. There exist $T > 0$ and $\varepsilon^* > 0$, such that, for all $\varepsilon \in]0, \varepsilon^*]$, for every $\tilde{u}_0 \in B_{R/2}(u_0)$, the Cauchy problem*

$$\frac{d\tilde{u}^{\varepsilon}}{dt}(t) = \varepsilon g(t, \tilde{u}^{\varepsilon}(t)), \quad \tilde{u}^{\varepsilon}(0) = \tilde{u}_0 \tag{2.6}$$

has a unique solution in $C^1([0, T/\varepsilon], X)$ remaining in K_R for all $t \leq T/\varepsilon$. Furthermore, for $\tilde{u}_0 = u_0$, $u^{\varepsilon}(t)$ remains in K .

The main result of this chapter is now stated in the following theorem.

Theorem 2.13 *Suppose that Assumptions 2.8 and 2.11 are satisfied and let $\eta = (\nu + d + 1)^{-1}$. There exists $0 < \hat{\varepsilon} < \varepsilon^*$, $C_1 > 0$, $C_2 > 0$, $0 < \varrho < R/2$ such that: for all $0 < \varepsilon \leq \hat{\varepsilon}$, there exist a change of variable $\phi^\varepsilon \in \mathcal{A}_{W,R}^\omega$ and an autonomous vector field $\mathcal{G}^\varepsilon \in \mathbf{A}_R$ satisfying*

$$\forall \tilde{u}_0 \in B_\varrho(u_0), \forall t \in [0, T/\varepsilon], \quad \|u^\varepsilon(t) - \phi^\varepsilon(t, \psi^\varepsilon(t, \tilde{u}_0))\| \leq C_1 \exp(-C_2 \varepsilon^{-\eta}),$$

where ψ^ε is the t -flow of the autonomous equation $\frac{d}{dt}\psi^\varepsilon(t, \tilde{u}_0) = \varepsilon \mathcal{G}^\varepsilon(\psi^\varepsilon(t, \tilde{u}_0))$.

The remaining of the paper is concerned with the proof of previous theorem under (2.11).

3 Technical proofs and intermediate results

3.1 Preliminary Lemmas

Our next task is to justify that the operators Λ and Γ^ε are well-defined.

Lemma 3.1 *Let $0 < \tilde{\sigma} < \sigma \leq 2W$ and $0 < \tilde{\rho} < \rho \leq 2R$. If ϕ and $\tilde{\phi}$ are two functions of $\mathcal{A}_{\tilde{\sigma},\rho}^\omega$ satisfying*

$$\|\phi - \text{id}\|_{\sigma,\rho} \leq \frac{\tilde{\rho}}{2}, \quad \|\tilde{\phi} - \text{id}\|_{\sigma,\rho} \leq \frac{\tilde{\rho}}{2},$$

then the following statements hold

1. The function $\langle \partial_u \phi \rangle^{-1}$ is well-defined in $\mathbf{A}_{\rho-\tilde{\rho}}$ and we have

$$\|\langle \partial_u \phi \rangle^{-1}\|_{\rho-\tilde{\rho}} \leq 2.$$

2. The mappings $\Lambda(\phi)$ and $\Gamma^\varepsilon(\phi)$ belong to respectively $\mathcal{A}_{\sigma,\rho-\tilde{\rho}}^\omega$ and $\mathcal{A}_{\sigma-\tilde{\sigma},\rho-\tilde{\rho}}^\omega$ and

$$\|\Lambda(\phi)\|_{\sigma,\rho-\tilde{\rho}} \leq 4C, \quad \|\Gamma^\varepsilon(\phi) - \text{id}\|_{\sigma-\tilde{\sigma},\rho-\tilde{\rho}} \leq 4C\kappa_0 \tilde{\sigma}^{-(\nu+d)} |\varepsilon|. \quad (3.1)$$

3. The mappings $\Lambda(\phi)$ and $\Gamma^\varepsilon(\phi)$ satisfy the following estimates:

$$\|\Lambda(\phi) - \Lambda(\tilde{\phi})\|_{\sigma,\rho-\tilde{\rho}} \leq \frac{16C}{\tilde{\rho}} \|\phi - \tilde{\phi}\|_{\sigma,\rho}, \quad (3.2)$$

$$\|\Gamma^\varepsilon(\phi) - \Gamma^\varepsilon(\tilde{\phi})\|_{\sigma-\tilde{\sigma},\rho-\tilde{\rho}} \leq 16C\kappa_0 \tilde{\rho}^{-1} \tilde{\sigma}^{-(\nu+d)} |\varepsilon| \|\phi - \tilde{\phi}\|_{\sigma,\rho}. \quad (3.3)$$

Proof. Applying Cauchy's estimate to the function $\phi - \text{id}$, we obtain

$$\|\partial_u \phi - \text{Id}\|_{\sigma, \rho - \tilde{\rho}} \leq \frac{1}{\tilde{\rho}} \|\phi - \text{id}\|_{\sigma, \rho} \leq \frac{1}{2}.$$

On the one hand, this gives $\|\partial_u \phi\|_{\sigma, \rho - \tilde{\rho}} \leq 3/2$ and on the other hand, using Neumann series, we can assert that $\langle \partial_u \phi \rangle^{-1}$ is well-defined in $\mathbf{A}_{\rho - \tilde{\rho}}$ with

$$\|\langle \partial_u \phi \rangle^{-1}\|_{\rho - \tilde{\rho}} \leq \sum_{k \geq 1} \|\langle \partial_u \phi \rangle - \text{Id}\|_{\rho - \tilde{\rho}}^k \leq \sum_{k \geq 1} \|\partial_u \phi - \text{Id}\|_{\sigma, \rho - \tilde{\rho}}^k \leq 2.$$

This proves the first statement. Now, consider the functions G and Φ associated with the quasi-periodic functions g and ϕ , such that $g(t, u) = G(t\omega, u)$ and $\phi(t, u) = \Phi(t\omega, u)$. Then the composition $(g \circ \phi)(t, u) = g(t, \phi(t, u)) = G(t\omega, \Phi(t\omega, u))$ belongs to $\mathcal{A}_{\sigma, \rho - \tilde{\rho}}^\omega$: as a matter of fact, for all $\theta \in D_\sigma$ and $u = u_1 + u_2 \in K_{\rho - \tilde{\rho}}$ with u_1 in K and $\|u_2\| \leq \rho - \tilde{\rho}$, the following inequality holds

$$\|\Phi(\theta, u) - u_1\| \leq \|\phi - \text{id}\|_{\sigma, \rho - \tilde{\rho}} + \|u_2\| \leq \frac{\tilde{\rho}}{2} + (\rho - \tilde{\rho}) \leq \rho - \frac{\tilde{\rho}}{2} \leq 2R, \quad (3.4)$$

so that, for all $\theta \in D_\sigma$, $\Phi(\theta, K_{\rho - \tilde{\rho}}) \subset K_{2R}$, and eventually

$$\|g \circ \phi\|_{\sigma, \rho - \tilde{\rho}} \leq \sup_{\theta \in D_\sigma} \|G(\theta)\|_{2R} \leq C.$$

In passing, we get $\|\langle g \circ \phi \rangle\|_{\rho - \tilde{\rho}} \leq C$. Collecting all terms in the expression of Λ it follows that $\|\Lambda(\phi)\|_{\sigma, \rho - \tilde{\rho}} \leq 4C$. Using Lemma 2.10 and the fact that $\langle \Lambda(\phi) \rangle = 0$, we also have

$$\|\Gamma^\varepsilon(\phi) - \text{id}\|_{\sigma - \tilde{\sigma}, \rho - \tilde{\rho}} \leq 4C\kappa_0 \tilde{\sigma}^{-(\nu+d)} |\varepsilon|.$$

In order to prove the last statement, we denote, for the sake of brevity

$$\begin{aligned} A_1 &= \left\| g \circ \phi - g \circ \tilde{\phi} \right\|_{\sigma, \rho - \tilde{\rho}} \\ A_2 &= \left\| \partial_u \phi - \partial_u \tilde{\phi} \right\|_{\sigma, \rho - \tilde{\rho}} \left\| \langle \partial_u \phi \rangle^{-1} \right\|_{\rho - \tilde{\rho}} \left\| \langle g \circ \phi \rangle \right\|_{\rho - \tilde{\rho}} \\ A_3 &= \left\| \partial_u \tilde{\phi} \right\|_{\sigma, \rho - \tilde{\rho}} \left\| \langle \partial_u \phi \rangle^{-1} - \langle \partial_u \tilde{\phi} \rangle^{-1} \right\|_{\rho - \tilde{\rho}} \left\| \langle g \circ \phi \rangle \right\|_{\rho - \tilde{\rho}} \\ A_4 &= \left\| \partial_u \tilde{\phi} \right\|_{\sigma, \rho - \tilde{\rho}} \left\| \langle \partial_u \tilde{\phi} \rangle^{-1} \right\|_{\rho - \tilde{\rho}} \left\| \langle g \circ \phi - g \circ \tilde{\phi} \rangle \right\|_{\rho - \tilde{\rho}} \end{aligned}$$

so that we have $\left\| \Lambda(\phi) - \Lambda(\tilde{\phi}) \right\|_{\sigma, \rho - \tilde{\rho}} \leq A_1 + A_2 + A_3 + A_4$. First of all, taking (3.4) into account

$$A_1 \leq \|\partial_u g\|_{\sigma, \rho - \tilde{\rho}/2} \left\| \phi - \tilde{\phi} \right\|_{\sigma, \rho - \tilde{\rho}} \leq \frac{2C}{\tilde{\rho}} \left\| \phi - \tilde{\phi} \right\|_{\sigma, \rho}.$$

We then have $\|\partial_u \phi\|_{\sigma, \rho-\tilde{\rho}} \leq 3/2$, $\|\langle \partial_u \phi \rangle^{-1}\|_{\rho-\tilde{\rho}} \leq 2$, $\|\langle g \circ \phi \rangle\|_{\rho-\tilde{\rho}} \leq C$ and similarly for $\tilde{\phi}$. It follows that

$$A_2 \leq \frac{2C}{\tilde{\rho}} \|\phi - \tilde{\phi}\|_{\sigma, \rho} \quad \text{and} \quad A_4 \leq 3A_1 \leq \frac{6C}{\tilde{\rho}} \|\phi - \tilde{\phi}\|_{\sigma, \rho}.$$

Regarding A_3 , we notice that $\langle \partial_u \phi \rangle^{-1} - \langle \partial_u \tilde{\phi} \rangle^{-1} = \langle \partial_u \phi \rangle^{-1} \langle \partial_u \tilde{\phi} - \partial_u \phi \rangle \langle \partial_u \tilde{\phi} \rangle^{-1}$ to obtain

$$\begin{aligned} \left\| \langle \partial_u \phi \rangle^{-1} - \langle \partial_u \tilde{\phi} \rangle^{-1} \right\|_{\rho-\tilde{\rho}} &\leq \left\| \langle \partial_u \phi \rangle^{-1} \right\|_{\rho-\tilde{\rho}} \left\| \langle \partial_u \tilde{\phi} \rangle^{-1} \right\|_{\rho-\tilde{\rho}} \left\| \partial_u \tilde{\phi} - \partial_u \phi \right\|_{\sigma, \rho-\tilde{\rho}} \\ &\leq \frac{4}{\tilde{\rho}} \|\phi - \tilde{\phi}\|_{\sigma, \rho}. \end{aligned}$$

Therefore, $A_3 \leq (6C/\tilde{\rho}) \|\phi - \tilde{\phi}\|_{\sigma, \rho}$, and this completes the proof. Estimate (3.3) is obtained by application of Lemma 2.10. $\blacksquare$

To obtain a quasi-solution, we fix $n \in \mathbb{N}^*$ and we consider the sequence of functions

$$\forall t \in \mathbb{R}, \phi^{[0]}(t) \equiv \text{id}_{X_C} \in \mathbf{A}_{2R}, \quad \phi^{[k+1]} = \Gamma^\varepsilon(\phi^{[k]}), \quad k = 0, \dots, n, \quad (3.5)$$

and

$$\mathcal{G}^{[k]} = \langle \partial_u \phi^{[k]} \rangle^{-1} \langle g \circ \phi^{[k]} \rangle \quad k = 0, \dots, n. \quad (3.6)$$

We denote by $r_n = R/(n+1)$, $w_n = W/(n+1)$ and for all $k = 0, \dots, n+1$, $R_k = 2R - kr_n$, $W_k = 2W - kw_n$ so that $R_0 = 2R$, $W_0 = 2W$, $R_{n+1} = R$ and $W_{n+1} = W$. We first show that the two sequences of $\phi^{[k]}$'s and $\mathcal{G}^{[k]}$'s are well-defined respectively in $\mathcal{A}_{W_k, R_k}^\omega$ and $\mathbf{A}_{R_{k+1}}$.

Lemma 3.2 *Let $\kappa_1 = \frac{16C\kappa_0}{W^{(\nu+d)}R}$. For $\varepsilon \in \mathbb{C}$ such that $(n+1)^{\nu+d+1}|\varepsilon| \leq 2(\kappa_1)^{-1}$, the maps $\phi^{[k]}$, for $k = 0, \dots, n+1$, and the vector fields $\mathcal{G}^{[k]}$, for $k = 0, \dots, n$, are well-defined respectively in $\mathcal{A}_{W_k, R_k}^\omega$ and $\mathbf{A}_{R_{k+1}}$. Moreover, we have*

$$\|\phi^{[k]} - \text{id}\|_{W_k, R_k} \leq \frac{r_n}{2} \quad \text{and} \quad \|\mathcal{G}^{[k]}\|_{R_{k+1}} \leq 2C. \quad (3.7)$$

Proof. The proof proceeds by induction. The case $k = 0$ is clear since $\phi^{[0]}(t) = \text{id}_{X_C}$ and $\mathcal{G}^{[0]} = \langle g \rangle$. Now, assuming that $\phi^{[k]} \in \mathcal{A}_{W_k, R_k}^\omega$ and that (3.7) is fulfilled, the function $\phi^{[k+1]} = \Gamma^\varepsilon(\phi^{[k]})$ is, by Lemma 3.1 with $\tilde{\sigma} = w_n$ and $\tilde{\rho} = r_n$, well-defined in $\mathcal{A}_{W_{k+1}, R_{k+1}}^\omega$. From inequality (3.1), we further have

$$\|\phi^{[k+1]} - \text{id}\|_{W_{k+1}, R_{k+1}} \leq 4C\kappa_0 w_n^{-(\nu+d)} |\varepsilon| \leq \frac{r_n}{2},$$

and repeating some of the arguments of Lemma 3.1, we get, if $k \leq n-1$, that

$$\|\langle \partial_u \phi^{[k+1]} \rangle^{-1}\|_{R_{k+2}} \leq 2 \quad \text{and} \quad \|\langle g \circ \phi^{[k+1]} \rangle\|_{R_{k+2}} \leq C$$

thus proving the statement for $k+1$. $\blacksquare$

3.2 Resolution of (1.11) with polynomial remainder estimates

So as to estimate the remainder in equation (1.11), we introduce $\xi^{[0]} = -\varepsilon\Lambda(\text{id})$ and the sequence of defects $\xi^{[k]}$ for $k = 1, \dots, n$, defined by the relation

$$\xi^{[k]} = \varepsilon\Lambda(\phi^{[k-1]}) - \varepsilon\Lambda(\phi^{[k]}) \in \mathcal{A}_{W_k, R_{k+1}}^\omega. \quad (3.8)$$

We now show that $\xi^{[n]}$ is bounded by ε^{n+1} .

Lemma 3.3 (Existence of quasi-solutions to (1.11)) *Let κ_1 as in Lemma 3.2. Given $n \geq 0$, consider the sequences $(\phi^{[k]})_{k=0, \dots, n+1}$, $(\mathcal{G}^{[k]})_{k=0, \dots, n}$ and $(\xi^{[k]})_{k=0, \dots, n}$, respectively defined by (3.5), (3.6) and (3.8). Then, for $(n+1)^{\nu+d+1}|\varepsilon| < 2(\kappa_1)^{-1}$, the following statements hold true:*

1. *The mappings $\phi^{[n]}$, $\mathcal{G}^{[n]}$ and $\xi^{[n]}$ are well-defined and belong to $\mathcal{A}_{W+w_n, R+r_n}^\omega$, $\mathbf{A}_R$ and $\mathcal{A}_{W_n, R}^\omega$.*
2. *The mappings $\phi^{[n]} - \text{id}$, $\mathcal{G}^{[n]}$ and $\xi^{[n]}$ are bounded as follows*

$$\|\phi^{[n]} - \text{id}\|_{W+w_n, R+r_n} \leq \frac{r_n}{2}, \quad (3.9)$$

$$\|\mathcal{G}^{[n]}\|_R \leq 2C, \quad (3.10)$$

$$\|\xi^{[n]}\|_{W_n, R} \leq 4C|\varepsilon|(\kappa_1(n+1)^{\nu+d+1}|\varepsilon|)^n. \quad (3.11)$$

3. *For all $t \in \mathbb{R}$, the mapping $\phi^{[n]}(t) \in \mathbf{A}_R$ is a one-to-one map from K_R onto $\phi^{[n]}(t)(K_R)$. Its inverse is defined at least on K_R with values in K_{R+r_n} and is analytic. Moreover, $(\phi^{[n]}(t))^{-1}(K_{R-r_n}) \subset K_R$*

Proof of Lemma 3.3 The first point of this Lemma and estimates (3.9) and (3.10) are direct consequences of Lemma 3.2 for $k = n$. Then, using the fact that

$$\xi^{[n]} = \varepsilon\left(\Lambda(\phi^{[n-1]}) - \Lambda(\phi^{[n]})\right),$$

Lemma 3.1 with $\tilde{\sigma} = w_n$ and $\tilde{\rho} = r_n$ leads to

$$\begin{aligned} \|\xi^{[n]}\|_{W_n, R} &\leq \frac{16C|\varepsilon|}{r_n} \|\phi^{[n-1]} - \phi^{[n]}\|_{W_n, R+r_n} \\ &\leq \frac{16C|\varepsilon|}{r_n} \|\Gamma^\varepsilon(\phi^{[n-2]}) - \Gamma^\varepsilon(\phi^{[n-1]})\|_{W_n, R+r_n} \\ &\leq \frac{16C|\varepsilon|}{r_n} 16C\kappa_0 w_n^{-(\nu+d)} r_n^{-1} |\varepsilon| \|\phi^{[n-2]} - \phi^{[n-1]}\|_{W_{n-1}, R_{n-1}} \\ &\leq \dots \\ &\leq \frac{16C|\varepsilon|}{r_n} (16C\kappa_0 w_n^{-(\nu+d)} r_n^{-1} |\varepsilon|)^{n-1} \|\text{id} - \phi^{[1]}\|_{W_1, R_1}. \end{aligned}$$

Besides, $\|\phi^{[1]} - \text{id}\|_{W_1, R_1} = \|\Gamma^\varepsilon(\phi^{[0]}) - \text{id}\|_{W_1, R_1} \leq 4C\kappa_0 w_n^{-(\nu+d)}|\varepsilon|$ and so

$$\|\xi^{[n]}\|_{W_n, R} \leq \frac{16C|\varepsilon|}{r_n} (16C\kappa_0 w_n^{-(\nu+d)} r_n^{-1} |\varepsilon|)^{n-1} 4C\kappa_0 w_n^{-(\nu+d)} |\varepsilon|.$$

This proves (3.11). As far as the last statement is concerned, let $t \in \mathbb{R}$ be given and consider $u, \tilde{u} \in K_R$ such that $\phi^{[n]}(t, u) = \phi^{[n]}(t, \tilde{u})$. We have

$$\begin{aligned} \|u - \tilde{u}\| &\leq \|\partial_u \phi^{[n]} - \text{id}\|_{W_n, R_{n+1}} \|u - \tilde{u}\| \\ &\leq \frac{1}{r_n} \|\phi^{[n]} - \text{id}\|_{W_n, R_n} \|u - \tilde{u}\| \leq \frac{1}{2} \|u - \tilde{u}\| \end{aligned}$$

This proves that $\phi^{[n]}(t)$ is a one-to-one map from K_R to its image. It remains to show that this image is included in K_{R+r_n} . Let $u = v + \tilde{v}$ be in K_R with $v \in K$ and $\alpha := \|\tilde{v}\| < R$. We consider the sequence u_k defined by

$$u_0 = u, \quad u_{k+1} = u_k - \phi^{[n]}(t, u_k) + u.$$

First of all, whenever $u_k \in K_{(R+\alpha+r_n)/2}$, noticing that $(R + \alpha + r_n)/2 < R + r_n$ and using the second statement, we have

$$\|u_{k+1} - v\| \leq \|\tilde{v}\| + \|\phi^{[n]} - \text{id}\|_{W_n, R_n} < \frac{R + \alpha + r_n}{2}$$

and, by induction, the sequence $(u_k)_{k \in \mathbb{N}}$ belongs to $K_{(R+\alpha+r_n)/2}$. Besides, we have

$$\begin{aligned} \|u_{k+1} - u_k\| &= \|(\phi^{[n]}(t, u_k) - u_k) - (\phi^{[n]}(t, u_{k-1}) - u_{k-1})\| \\ &\leq \|\partial_u \phi^{[n]} - \text{id}\|_{W_n, R_n - (R+r_n-\alpha)/2} \|u_k - u_{k-1}\| \\ &\leq \frac{r_n}{R + r_n - \alpha} \|u_k - u_{k-1}\|. \end{aligned}$$

Since $\alpha < R$, $(u_k)_{k \in \mathbb{N}}$ is a Cauchy sequence of the complete space $X_{\mathbb{C}}$ and thus converges. Its limit lies in K_{R+r_n} for $\alpha < R$ and in K_R if $\alpha < R - r_n$. Finally, the Inverse Function Theorem shows that $(\phi^{[n]}(t))^{-1}$ is analytic. $\blacksquare$

Remark 3.4 We note that the map $\xi^{[n]}$ satisfies the following bound equivalent to (3.11)

$$\frac{\kappa_0}{w_n^{\nu+d}} \|\xi^{[n]}\|_{W_n, R} \leq \frac{R}{4} (\kappa_1(n+1)^{\nu+d+1} |\varepsilon|)^{n+1}.$$

Lemma 3.5 Let κ_1 be the constant of Lemma 3.3 and let $n \in \mathbb{N}^*$ be given. Consider $\phi^{[n]} \in \mathcal{A}_{W_n, R+r_n}^\omega$ the quasi-solution of (1.11) obtained in Lemma 3.3, $\phi^* \in \mathcal{A}_{W_n, R+r_n}^\omega$ such that

$$\|\phi^* - \text{id}\|_{W_n, R+r_n} \leq \frac{r_n}{2}$$

and the defect ξ^* be defined as

$$(t, u) \rightarrow \partial_t \phi^*(t, u) - \varepsilon \Lambda(\phi^*)(t, u) \in \mathcal{A}_{W,R}^\omega.$$

Finally, we assume that ξ^* satisfies

$$\frac{\kappa_0}{w_n^{\nu+d}} \|\xi^*\|_{W,R} \leq \tilde{C} |\varepsilon|^{n+1}$$

where $\tilde{C}$ is a constant independent of ε . Then for all $(n+1)^{\nu+d+1} |\varepsilon| \leq (2\kappa_1)^{-1}$,

$$\|\phi^{[n]} - \phi^*\|_{\frac{w_n}{2}, r_n} \leq 2^{\nu+d} \left(\frac{R}{4} (\kappa_1(n+1)^{\nu+d+1} |\varepsilon|)^{n+1} + 2\tilde{C} |\varepsilon|^{n+1} \right). \quad (3.12)$$

Proof. Using Lemma 2.10 and 3.1, first with $\tilde{\sigma} = w_n/2$ and $\tilde{\rho} = r_n$ and then with $\tilde{\sigma} = w_n$ and $\tilde{\rho} = r_n$ leads to

$$\begin{aligned} \|\phi^{[n]} - \phi^*\|_{\frac{w_n}{2}, r_n} &\leq 2^{\nu+d} \frac{\kappa_0}{w_n^{\nu+d}} \|\partial_t \phi^{[n]} - \partial_t \phi^*\|_{w_n, r_n} \\ &\leq 2^{\nu+d} \frac{\kappa_0}{w_n^{\nu+d}} \|\varepsilon \Lambda(\phi^{[n-1]}) - \varepsilon \Lambda(\phi^*) - \xi^*\|_{w_n, r_n} \\ &\leq 2^{\nu+d} \left(\frac{16C\kappa_0}{w_n^{\nu+d} r_n} |\varepsilon| \|\phi^{[n-1]} - \phi^*\|_{w_n, 2r_n} + \frac{\kappa_0}{w_n^{\nu+d}} \|\xi^*\|_{w_n, r_n} \right) \\ &\leq 2^{\nu+d} \left(\kappa_1(n+1)^{\nu+d+1} |\varepsilon| \|\phi^{[n-1]} - \phi^*\|_{w_n, 2r_n} + \tilde{C} |\varepsilon|^{n+1} \right) \\ &\leq \dots \\ &\leq 2^{\nu+d} \left((\kappa_1(n+1)^{\nu+d+1} |\varepsilon|)^n \|\text{id} - \phi^*\|_{W-w_n, R} \right. \\ &\quad \left. + \sum_{k=0}^{n-1} (\kappa_1(n+1)^{\nu+d+1} |\varepsilon|)^k \tilde{C} |\varepsilon|^{n+1} \right) \\ &\leq 2^{\nu+d} \left((\kappa_1(n+1)^{\nu+d+1} |\varepsilon|)^n \frac{\kappa_0}{w_n^{\nu+d}} |\varepsilon| \|\Lambda(\phi^*)\|_{W,R} + 2\tilde{C} |\varepsilon|^{n+1} \right). \end{aligned}$$

The result stems from

$$\frac{\kappa_0}{w_n^{\nu+d}} \|\Lambda(\phi^*)\|_{W_n, R} \leq 4C \frac{\kappa_0}{w_n^{\nu+d}} \leq \kappa_1(n+1)^{\nu+d+1} R/4.$$

■

3.3 Proof of Theorem 2.13

In this section, we derive the main theorem, namely an averaging result for problems under the form (1.1). Exponential error estimates are obtained under the condition that g is analytic with respect to both variables.

Lemma 3.6 Define $\eta = (\nu + d + 1)^{-1}$ and $\kappa_1 = \frac{16C\kappa_0}{W^{\nu+d}R}$. Let $\kappa \geq \kappa_1$ be given and suppose $0 < |\varepsilon| < (e\kappa)^{-1}$. Finally, let us define $n_\varepsilon := \lfloor (e\kappa|\varepsilon|)^{-\eta} \rfloor - 1$. Then one has $n_\varepsilon \geq 0$ and estimate (3.11) for $n = n_\varepsilon$ becomes either

$$\|\xi^{[n]}\|_{W_{n,R}} \leq 4C e\kappa^{-1} e^{-(e\kappa|\varepsilon|)^{-\eta}}$$

or

$$\frac{\kappa_0}{w_n^{\nu+d}} \|\xi^{[n]}\|_{W_{n,R}} \leq \frac{Re}{4} e^{-(e\kappa|\varepsilon|)^{-\eta}}.$$

Moreover, condition $(n+1)^{\nu+d+1}|\varepsilon| < 2\kappa_1^{-1}$ of Lemma 3.3 for $n = n_\varepsilon$ is satisfied.

Proof. By definition of n_ε , $(e\kappa|\varepsilon|)^{-\eta} - 1 < (n_\varepsilon + 1) \leq (e\kappa|\varepsilon|)^{-\eta} \leq (e\kappa_1|\varepsilon|)^{-\eta}$. Choosing $n = n_\varepsilon$ in estimate (3.11) thus leads to

$$\|\xi^{[n_\varepsilon]}\|_{W_{n,R}} \leq 4C |\varepsilon| e^{-n_\varepsilon} \leq 4C |\varepsilon| e^{-(e\kappa|\varepsilon|)^{-\eta}+2} \leq 4C e\kappa^{-1} e^{-(e\kappa|\varepsilon|)^{-\eta}}$$

and to

$$\frac{\kappa_0}{w_n^{\nu+d}} \|\xi^{[n]}\|_{W_{n,R}} \leq \frac{4C\kappa_0}{\kappa_1 W^{\nu+d}} (\kappa_1(n_\varepsilon + 1)^{\nu+d+1}|\varepsilon|)^{n_\varepsilon+1} \leq \frac{R}{4} e^{-(n_\varepsilon+1)} \leq \frac{Re}{4} e^{-(e\kappa|\varepsilon|)^{-\eta}}.$$

Since $|\varepsilon| < (e\kappa)^{-1}$, we further have $(e\kappa|\varepsilon|)^{-\eta} > 1$ and $n_\varepsilon \geq 0$. Finally, $(n_\varepsilon + 1)^{\nu+d+1}|\varepsilon| \leq (e\kappa_1)^{-1} < 2\kappa_1^{-1}$. $\blacksquare$

Proposition 3.7 Suppose that Assumptions 2.8 and 2.11 are satisfied, let $\kappa \geq \frac{16C\kappa_0}{W^{\nu+d}R}$ be given and

$$\tilde{\varepsilon} = \min \left(\varepsilon^*, (e\kappa)^{-1}, \left(\frac{R}{2TC} \right)^{\frac{1}{\eta}} (e\kappa)^{-1}, \frac{R}{8Ce^2 m! 2^m} \right),$$

where $m = \lfloor \nu + d + 2 \rfloor$. Consider $0 < \varepsilon < \tilde{\varepsilon}$ and $n_\varepsilon = \lfloor (e\kappa|\varepsilon|)^{-\eta} \rfloor - 1$. Denoting $\phi^\varepsilon := \phi^{[n_\varepsilon]}$,

$$\mathcal{G}^\varepsilon := \langle \partial_u \phi^\varepsilon \rangle^{-1} \langle g \circ \phi^\varepsilon \rangle \quad (3.13)$$

and $\psi^\varepsilon(t)$ the time-flow of the autonomous equation with vector field $\mathcal{G}^\varepsilon$

$$\frac{d}{dt} \psi^\varepsilon(t, u_0) = \varepsilon \mathcal{G}^\varepsilon(\psi^\varepsilon(t, u_0))$$

one can assert that

$$\forall t \in [0, T/\varepsilon], \quad \|u^\varepsilon(t) - \phi^\varepsilon(t, \psi^\varepsilon(t, u_0))\| \leq 4C e\kappa^{-1} \exp \left(-\frac{1}{2} (e\kappa\varepsilon)^{-\eta} \right), \quad (3.14)$$

where $\eta = 1/(\nu + d + 1)$.

Proof. According to Lemma 3.6, we can apply Lemma 3.3 with $n = n_\varepsilon$ and obtain the following bounds for ϕ^ε and $\xi^\varepsilon := \xi^{[n_\varepsilon]}$:

$$\|\phi^\varepsilon - \text{id}\|_{W+w_{n_\varepsilon}, R+r_{n_\varepsilon}} \leq \frac{r_{n_\varepsilon}}{2}, \quad \|\mathcal{G}^\varepsilon\|_R \leq 2C \quad \text{and} \quad \|\xi^\varepsilon\|_{W+w_{n_\varepsilon}, R} \leq 4C e \kappa^{-1} e^{-(e\kappa|\varepsilon|)^{-\eta}}.$$

The vector field $\mathcal{G}^\varepsilon$ is analytic on K_R thus Lipschitz-continuous on K_ρ for all $0 < \rho < R$. Hence, there exists $T_1 > 0$ such that the map $\psi^\varepsilon(t, u_0)$ is well-defined and remains in K_R for all $t \in [0, T_1/\varepsilon]$. By differentiating $v^\varepsilon(t) := \phi^\varepsilon(t, \psi^\varepsilon(t, u_0))$, we obtain

$$\frac{d}{dt}v^\varepsilon(t) = \partial_t \phi^\varepsilon(t, \psi^\varepsilon(t, u_0)) + \varepsilon \partial_u \phi^\varepsilon(t, \psi^\varepsilon(t, u_0)) \mathcal{G}^\varepsilon(\psi^\varepsilon(t, u_0)).$$

Since, for all $v \in K_{R-r_{n_\varepsilon}}$,

$$\partial_t \phi^\varepsilon(t, v) = \varepsilon \Lambda(\phi^\varepsilon)(t, v) + \xi^\varepsilon(t, v) = \varepsilon g(t, \phi^\varepsilon(t, v)) - \varepsilon \partial_u \phi^\varepsilon(t, v) \mathcal{G}^\varepsilon(v) + \xi^\varepsilon(t, v),$$

we have for $t \in [0, T_1/\varepsilon]$

$$\frac{d}{dt}v^\varepsilon(t) = \varepsilon g(t, v^\varepsilon(t)) + \xi^\varepsilon(t, (\phi^\varepsilon)^{-1}(t, v^\varepsilon(t))).$$

Since $0 < \varepsilon < \varepsilon^*$, Theorem 2.12 ensures that $u^\varepsilon(t) \in K$ for all $t \in [0, T/\varepsilon]$, hence we can consider $u^\varepsilon(t)$ and $v^\varepsilon(t)$ on the interval $[0, \min(T, T_1)/\varepsilon]$. Moreover, as long as $v^\varepsilon(t)$ remains in $K_{R-r_{n_\varepsilon}}$, we have $(\phi^\varepsilon)^{-1}(t, v^\varepsilon(t))$ in K_R (Lemma 3.3) and

$$\begin{aligned} \left\| \frac{d}{dt}u^\varepsilon(t) - \frac{d}{dt}v^\varepsilon(t) \right\| &\leq \varepsilon \|g(t, u^\varepsilon(t)) - g(t, v^\varepsilon(t))\| + \|\xi^\varepsilon(t, (\phi^\varepsilon)^{-1}(t, v^\varepsilon(t)))\| \\ &\leq \frac{C\varepsilon}{R} \|u^\varepsilon(t) - v^\varepsilon(t)\| + \|\xi^\varepsilon\|_{W+w_{n_\varepsilon}, R}, \end{aligned}$$

where we have used that $g(t)$, being analytic on K_{2R} and bounded by C , is C/R -Lipschitz continuous on K_R . Owing to Grönwall Lemma, we thus obtain the following bound

$$\|u^\varepsilon(t) - v^\varepsilon(t)\| \leq \|\xi^\varepsilon\|_{W+w_{n_\varepsilon}, R} e^{\frac{Ct\varepsilon}{R}} \leq 4C e \kappa^{-1} \exp\left(CR^{-1}t|\varepsilon| - (e\kappa|\varepsilon|)^{-\eta}\right).$$

Since $\varepsilon < \varepsilon_2 := \left(\frac{RT^{-1}}{2C(e\kappa)^\eta}\right)^{\frac{1}{\eta}}$, we have for all $t \in [0, T/\varepsilon]$

$$\exp\left(CR^{-1}t|\varepsilon| - (e\kappa|\varepsilon|)^{-\eta}\right) \leq \exp\left(CR^{-1}T - (e\kappa|\varepsilon|)^{-\eta}\right) \leq \exp\left(-\frac{1}{2}(e\kappa|\varepsilon|)^{-\eta}\right).$$

Using the inequality $e^{-x} \leq \frac{m!}{x^m}$ valid for all $x > 0$, we have, since $m\eta \geq 1$ and $e\kappa|\varepsilon| \leq 1$

$$\exp\left(-\frac{1}{2}(e\kappa|\varepsilon|)^{-\eta}\right) \leq m! 2^m (e\kappa|\varepsilon|)^{m\eta} \leq m! 2^m e\kappa|\varepsilon|$$

Now, since $\varepsilon < \min(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ with $\varepsilon_3 = R(8Ce^2m!2^m)^{-1}$, we further have

$$4Ce\kappa^{-1} \exp\left(CR^{-1}t|\varepsilon| - (e\kappa|\varepsilon|)^{-\eta}\right) \leq 4Ce\kappa^{-1} \exp\left(-\frac{1}{2}(e\kappa|\varepsilon|)^{-\eta}\right) \leq \frac{R}{2}.$$

Finally, by a bootstrap argument, $v^\varepsilon(t)$ exists and remains in K_R for all $t \in [0, T/\varepsilon]$ since,

$$\|u^\varepsilon(t) - v^\varepsilon(t)\| \leq 4C(e\kappa)^{-1}e^{-(1/2)(e\kappa|\varepsilon|)^{-\eta}} \leq \frac{R}{2} \leq R - r_{n_\varepsilon}.$$

Finally, inequality (3.14) holds for $t \in [0, T/\varepsilon]$. $\blacksquare$

Remark 3.8 *With the same proof, we have, for all $n \in \mathbb{N}, t \in [0, T/\varepsilon]$, a constant $C(n)$ independent of ε such that*

$$\|u^\varepsilon(t) - \phi^{[n]}(t, \psi^{[n]}(t, u_0))\| \leq C(n)\varepsilon^{n+1}.$$

One can ask what happens to Theorem 3.7 when we change the initial condition in equation (1.1). It turns out that the theorem remains valid under some conditions.

Proposition 3.9 *Let $\varrho = e^{-CTR^{-1}\frac{R}{4}}$ and $\tilde{u}_0 \in B_\varrho(u_0)$. Then, for all $0 < \varepsilon < \tilde{\varepsilon}$, the solution $\tilde{u}^\varepsilon$ of equation (2.6), the functions ϕ^ε and ψ^ε defined in Proposition 3.7 satisfy*

$$\forall t \in [0, T/\varepsilon], \quad \psi^\varepsilon(t, \tilde{u}_0) \in K_R \quad (3.15)$$

$$\forall t \in [0, T/\varepsilon], \quad \|u^\varepsilon(t, \tilde{u}_0) - \phi^\varepsilon(t, \psi^\varepsilon(t, \tilde{u}_0))\| \leq 4Ce\kappa^{-1} \exp\left(-\frac{1}{2}(e\kappa\varepsilon)^{-\eta}\right). \quad (3.16)$$

Proof. We adapt the proof of Proposition 3.7 with $\tilde{u}_0 \in B_\varrho(u_0)$. The vector field $\mathcal{G}^\varepsilon$ is analytic, hence there exists $T_2(\tilde{u}_0) > 0$ such that the map $\psi^\varepsilon(t, \tilde{u}_0)$ is well-defined and remains in K_R for all $t \in [0, \min(T, T_2(\tilde{u}_0))/\varepsilon]$. The map $\tilde{v}^\varepsilon(t) := \phi^\varepsilon(t, \psi^\varepsilon(t, \tilde{u}_0))$ is then well-defined and satisfies (following the proof of Theorem 2.12),

$$\|\partial_t \tilde{u}^\varepsilon(t) - \partial_t \tilde{v}^\varepsilon(t)\| \leq \frac{C\varepsilon}{R} \|\tilde{u}^\varepsilon(t) - \tilde{v}^\varepsilon(t)\| + \|\xi^\varepsilon\|_{W+w_{n_\varepsilon}, R}.$$

The computations made in the proof of Theorem 2.12 remains valid for

$$t \in [0, \min(T, T_2(\tilde{u}_0))/\varepsilon]$$

and we have

$$\forall t \in [0, T/\varepsilon], \quad \|\tilde{u}^\varepsilon(t) - \tilde{v}^\varepsilon(t)\| \leq 4C(e\kappa)^{-1} \exp(-C_2\varepsilon^{-\eta}) \leq \frac{R}{2}.$$

On the other hand, since $\tilde{u}_0 \in K_{R/2}$ a Grönwall Lemma gives

$$\forall t \in [0, T/\varepsilon], \quad \|\tilde{u}^\varepsilon(t) - u^\varepsilon(t)\| \leq e^{CTR^{-1}} \|\tilde{u}_0 - u_0\| \leq \frac{R}{4} \leq \frac{R}{2} - r_{n_\varepsilon}.$$

Hence, for all $t \in [0, T/\varepsilon]$, $\tilde{v}^\varepsilon(t)$ is in $K_{R-r_{n_\varepsilon}}$ and the proof is clear with a bootstrap argument. $\blacksquare$

Chapter 4

Quasi-stroboscopic procedure for a semi-linear Schrödinger equation

1 Introduction

In this chapter, we give a concrete example of quasi-stroboscopic averaging, namely a nonlinear Schrödinger equation which can be put under the form

$$\frac{d}{dt}u^\varepsilon = g(t, u^\varepsilon), \quad u^\varepsilon(0) = u_0. \quad (1.1)$$

Here, the function g is assumed to be quasi-periodic and analytic with respect to its two variables, which allows us to apply the Theorems of the previous chapter.

However, the analyticity in t for the function g is an issue for nonlinear waves equation or Schrödinger equation. The main problem is the unboundness of the Laplacian operator. In this work, this difficulty is overcome by introducing a finite dimensional space in which the Laplacian operator is bounded.

The main theorem of this chapter is then the application of the quasi-stroboscopic procedure to the projected Schrödinger equation (Theorem 4.2).

Given $d \geq 2$ an integer, and the d -dimensional torus $\mathbb{T}^d := [0, 1]^d$, we introduce $L^2_{\mathbb{C}}(\mathbb{T}^d)$ the Hilbert space of square integrable functions of $\mathbb{T}^d$ with values in $\mathbb{C}$ equipped with the usual scalar product $(\cdot, \cdot)$. The space $H^2_{per}([0, 1])$ denotes the Sobolev space of periodic functions on $[0, 1]$. Hence, for $j = 1, \dots, d$ we define the domain $\mathcal{D}_j := L^2_{\mathbb{C}}([0, 1]) \otimes \dots \otimes H^2_{per}([0, 1]) \otimes \dots \otimes L^2_{\mathbb{C}}([0, 1])$ of the linear operator

$$\Delta_j := \frac{\partial^2}{\partial x_j^2} : \mathcal{D}_j \subset L^2_{\mathbb{C}}(\mathbb{T}^d) \rightarrow L^2_{\mathbb{C}}(\mathbb{T}^d).$$

It is trivial that the operator Δ_j (for $j = 1, \dots, d$) satisfies the following statements.

- This is a self-adjoint operator (it means that for all $y_1, y_2 \in \mathcal{D}_j$ we have $(\Delta_j y_1, y_2) = (y_1, \Delta_j y_2)$).
- For all $i \in \llbracket 1, d \rrbracket$, the operator Δ_i commutes with the operator Δ_j .
- The spectrum of Δ_j is $2\pi\mathbb{N}$.

We are interested in the following nonlinear Schrödinger equation:

$$\begin{aligned} i\partial_t y^\varepsilon(t, x) &= - \sum_{j=1}^d \omega_j \Delta_j y^\varepsilon(t, x) + \varepsilon f(|y^\varepsilon(t, x)|^2) y^\varepsilon(t, x), \quad t \geq 0, \quad x \in \mathbb{T}^d \\ y^\varepsilon(0, \cdot) &= y_0. \end{aligned} \tag{1.2}$$

The function $f : \mathbb{R} \rightarrow \mathbb{R}$ is assumed to be real analytic and the vector of frequencies $\omega = (\omega_1, \dots, \omega_d)$ satisfies a strong non-resonance condition

Assumption 1.1 *There exist constants $\gamma > 0$, $\nu > 0$, such that*

$$|k \cdot \omega| \geq \gamma |k|^{-\nu},$$

for all k in $\mathbb{Z}^d$ and $k \neq 0$.

This assumption seems to restrict the choice of the ω_j 's. However, we recall that for $\nu > d - 1$, the set of vectors ω which do not satisfy Assumption 1.1 has Lebesgue measure 0 in $\mathbb{R}^d$. This assumption fixes the values of γ and ν throughout the remaining of this paper.

Notations In the sequel, we constantly use for $s > d/2$ the Sobolev space $H_{\mathbb{C}}^s(\mathbb{T}^d)$ defined by

$$H_{\mathbb{C}}^s(\mathbb{T}^d) = \{y \in L_{\mathbb{C}}^2(\mathbb{T}^d) \mid \forall \alpha \text{ s.t. } |\alpha| \leq s, \partial^\alpha y \in L_{\mathbb{C}}^2(\mathbb{T}^d)\}$$

equipped with the scalar product

$$\forall y, \tilde{y} \in H_{\mathbb{C}}^s(\mathbb{T}^d), \quad (y, \tilde{y})_s = ((1 - \Delta)^s y, (1 - \Delta)^s \tilde{y})$$

and the inner norm $\|\cdot\|_s$.

The condition $s > d/2$ implies that for any nonlinear function $g \in \mathcal{C}^s(\mathbb{C})$ satisfying $g(0) = 0$, there exists a constant C_g (depending on $\|y\|_\infty$ and $\|\tilde{y}\|_\infty$) (see [15]) such that

$$\|g(y)\|_s \leq C_g(\|y\|_\infty) \|y\|_s.$$

Besides, the space $H_{\mathbb{C}}^s(\mathbb{T}^d)$ is an algebra (for $y_1, y_2 \in H_{\mathbb{C}}^s(\mathbb{T}^d)$, we have $\|y_1 y_2\|_s \lesssim \|y_1\|_s \|y_2\|_s$).

We now give the following tame estimate

Lemma 1.2 *Let $s > d/2$ and consider a function $g \in \mathcal{C}^s(\mathbb{C}^2, \mathbb{C})$. Then for $M > 0$, there exists $C_g(M) > 0$ such that*

$$\|g(y, z) - g(\tilde{y}, z)\|_{L^2} \leq C_g(M) \|y - \tilde{y}\|_{L^2} \quad (1.3)$$

where $y, \tilde{y}, z \in H_{\mathbb{C}}^s(\mathbb{T}^d)$ satisfy $\|y\|_{\infty} \leq M$, $\|\tilde{y}\|_{\infty} \leq M$ and $\|z\|_{\infty} \leq M$.

Proof. We denote

$$C_g(M) = \sup_{|y_1| \leq M, |y_2| \leq M} |\partial_y g(y_1, y_2)|.$$

It follows that for all $x \in \mathbb{T}^D$, we have

$$|g(y(x), z(x)) - g(\tilde{y}(x), z(x))| \leq C_g(M) |y(x) - \tilde{y}(x)|,$$

which ends the proof. ■

Finally, we confirm that the Cauchy problem (1.2) is well-posed in our Hilbert setting.

Proposition 1.3 *Let $s > d/2$ and assume that $\tilde{y}_0 \in H_{\mathbb{C}}^s(\mathbb{T}^d)$ and $f \in \mathcal{C}^{s+1}(\mathbb{C})$. There exists $T > 0$, $1 \geq \varepsilon^* > 0$, such that, for all $\varepsilon \in]0; \varepsilon^*]$, the Cauchy problem (1.2) has a unique solution*

$$\tilde{y}^\varepsilon \in \mathcal{C}^0([0, T/\varepsilon], H_{\mathbb{C}}^s(\mathbb{T}^d)) \cap \mathcal{C}^1([0, T/\varepsilon], H_{\mathbb{C}}^{s-2}(\mathbb{T}^d))$$

and

$$\forall t \leq T/\varepsilon, \quad \|\tilde{y}^\varepsilon(t, \cdot)\|_s < 2 \|\tilde{y}_0\|_s.$$

Proposition 1.3 is proved in [13].

2 Projection of the nonlinear Schrödinger equation

We now write equation (1.2) under its Hamiltonian form. We then give the geometric properties satisfied by the Hamiltonian system. Next, we define a projection of this equation on a finite dimensional space. Then, we shall apply (in the next section) the quasi-stroboscopic procedure to the projected equation in the finite dimensional space.

2.1 Geometric form of the nonlinear Schrödinger equation

We denote, $\Re(y)$ and $\Im(y)$, respectively, the real and imaginary parts of $y \in \mathbb{C}$. We set for all $t \in [0, T/\varepsilon]$ and $x \in \mathbb{T}^d$,

$$u^\varepsilon(t, x) = \begin{pmatrix} \Im(\tilde{y}^\varepsilon(t, x)) \\ \Re(\tilde{y}^\varepsilon(t, x)) \end{pmatrix} \in \mathbb{R} \times \mathbb{R}. \quad (2.1)$$

Then, we introduce the real Hilbert spaces

$$X = H_{\mathbb{R}}^s(\mathbb{T}^d) \times H_{\mathbb{R}}^s(\mathbb{T}^d) \quad \text{and} \quad Z = L_{\mathbb{R}}^2(\mathbb{T}^d) \times L_{\mathbb{R}}^2(\mathbb{T}^d)$$

equipped with the scalar products

$$\forall u = \begin{pmatrix} p \\ q \end{pmatrix}, \tilde{u} = \begin{pmatrix} \tilde{p} \\ \tilde{q} \end{pmatrix} \in X, \quad (u, \tilde{u})_X = (p, \tilde{p})_s + (q, \tilde{q})_s + i[(p, \tilde{q})_s - (q, \tilde{p})_s],$$

$$\forall u = \begin{pmatrix} p \\ q \end{pmatrix}, \tilde{u} = \begin{pmatrix} \tilde{p} \\ \tilde{q} \end{pmatrix} \in Z, \quad (u, \tilde{u})_Z = (p, \tilde{p}) + (q, \tilde{q}) + i[(p, \tilde{q}) - (q, \tilde{p})]$$

and the inner norm denoted $\|\cdot\|_X$ and $\|\cdot\|_Z$. We note that for $u \in X$, the definition of the scalar product leads to $\|u\|_X = \|q + ip\|_s$.

For $j = 1, \dots, d$ and $s > d/2$, we denote

$$A_j = \begin{pmatrix} 0 & \Delta_j \\ -\Delta_j & 0 \end{pmatrix} \in \mathcal{L}(X, Z)$$

by A the collection of operators $(A_j)_{j=1, \dots, d}$ and

$$\omega \cdot A = \sum_{j=1}^d \omega_j A_j.$$

Finally, denoting for all $u = \begin{pmatrix} p \\ q \end{pmatrix} \in X$, $|u|_2^2 = p^2 + q^2$ equation (1.2) reads

$$\begin{aligned} \partial_t u^\varepsilon(t, x) &= (\omega \cdot A) u^\varepsilon(t, x) + \varepsilon h(u^\varepsilon(t, x)), \quad t \geq 0, x \in \mathbb{T}^d \\ u^\varepsilon(0, \cdot) &= u_0 := \begin{pmatrix} \Im(y_0) \\ \Re(y_0) \end{pmatrix} \in X, \end{aligned} \tag{2.2}$$

with

$$h : u \in X \rightarrow X \ni f(|u|_2^2) J^{-1} u, \quad \text{and} \quad J^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{2.3}$$

Proposition 2.1 *Let $s > d/2$, then the following statements hold true.*

1. *For all $j \in \llbracket 1, d \rrbracket$ and $u, \tilde{u} \in X$, we have*

$$(A_j u, \tilde{u})_Z = -(u, A_j \tilde{u})_Z.$$

2. *For all $j = 1, \dots, d$, the spectrum of A_j is $2i\pi\mathbb{Z}$.*

3. For $u_0 \in X$, we denote by $e^{t\omega_j A_j} u_0$, the solution of the linear equation

$$\frac{du}{dt} = \omega_j A_j u, \quad u(t=0) = u_0.$$

Then, the operator $e^{t\omega_j A_j}$ is 1-periodic with respect to $t\omega_j$ and is an isometry on Z and on X .

Proof. Let $u = \begin{pmatrix} p \\ q \end{pmatrix}, \tilde{u} = \begin{pmatrix} \tilde{p} \\ \tilde{q} \end{pmatrix} \in X$. We have that $p, q, \tilde{p}, \tilde{q}$ are in $\mathcal{D}_j$ and so $(\Delta_j p, \tilde{p})_s = (p, \Delta_j \tilde{p})_s$ and $(\Delta_j q, \tilde{q})_s = (q, \Delta_j \tilde{q})_s$. Then,

$$(A_j u, \tilde{u})_Z = (-\Delta_j p, \tilde{q})_s + (\Delta_j q, \tilde{p})_s + i[(\Delta_j q, \tilde{q})_s - (\Delta_j p, \tilde{p})_s] = -(u, A_j \tilde{u})_Z.$$

For $j = 1, \dots, d$ the spectrum of Δ_j is $2\pi\mathbb{N}$. Hence, the spectrum of A_j is $2i\pi\mathbb{Z}$.

Finally, since the operator A_j is skew-symmetric, using Stones Theorem, the operator $e^{t\omega_j A_j}$ is well-defined, it is an isometry on Z and on X and it is 1-periodic with respect to $t\omega_j$. $\blacksquare$

For $0 \leq t \leq T/\varepsilon$, we have $u^\varepsilon(t) := u^\varepsilon(t, \cdot)$ in X and more precisely, the following relation holds:

$$\forall t \in [0, T/\varepsilon], \quad \|u^\varepsilon(t)\|_X = \|\tilde{y}^\varepsilon(t)\|_s \leq 2 \|\tilde{y}_0\|_s = 2 \|u_0\|_X.$$

Introducing the set $K := \{u \in X, \quad \|u\|_X \leq 2 \|u_0\|_X\}$, it is clear that $u^\varepsilon(t)$ remains in K for all t in $[0, T/\varepsilon]$. We deduce from the tame estimate (1.3), the following proposition concerning the function h defined in (2.3)

Proposition 2.2 *Let $s > d/2$ and $f \in \mathcal{C}^1(\mathbb{R})$, then there exists a positive constant C_h such that,*

$$\forall u, \tilde{u} \in K, \quad \|h(u) - h(\tilde{u})\|_Z \leq C_h \|u - \tilde{u}\|_Z. \quad (2.4)$$

Proof. Let $u = \begin{pmatrix} p \\ q \end{pmatrix}, \tilde{u} = \begin{pmatrix} \tilde{p} \\ \tilde{q} \end{pmatrix} \in Z$, then $h(u) = f(p^2 + q^2) \begin{pmatrix} p \\ -q \end{pmatrix}$ and $h(\tilde{u}) = f(\tilde{p}^2 + \tilde{q}^2) \begin{pmatrix} \tilde{p} \\ -\tilde{q} \end{pmatrix}$. It follows that

$$\|h(u) - h(\tilde{u})\|_Z \leq \|f(p^2 + q^2)p - f(\tilde{p}^2 + \tilde{q}^2)\tilde{p}\|_{L^2} + \|f(p^2 + q^2)q - f(\tilde{p}^2 + \tilde{q}^2)\tilde{q}\|_{L^2}.$$

Denoting $g : (p, q) \rightarrow f(p^2 + q^2)p$, we have $g \in \mathcal{C}^1(\mathbb{C}^2, \mathbb{C})$, and

$$\|g(p, q) - g(\tilde{p}, \tilde{q})\|_{L^2} \leq \|g(p, q) - g(p, \tilde{q})\|_{L^2} + \|g(p, \tilde{q}) - g(\tilde{p}, \tilde{q})\|_{L^2}.$$

Then, since $\|p\|_\infty \leq \|u\|_X \leq 2 \|u_0\|_X$ (we have the same bound for $q, \tilde{p}, \tilde{q}$), there exists $C_g(2 \|u_0\|_X)$ (Lemma 1.2) such that

$$\|g(p, q) - g(\tilde{p}, \tilde{q})\|_{L^2} \leq 2C_g(2 \|u_0\|_X) \|u - \tilde{u}\|_Z.$$

Repeating the arguments for the second part of the inequality, the result is proved. $\blacksquare$

2.2 Hamiltonian character of equation (2.2)

In this context, a function $\tilde{g} : K \subset Z \rightarrow Z$ is said to be Hamiltonian if there exists a function $\mathcal{H} : K \rightarrow \mathbb{R}$ such that

$$\forall (u, v) \in K \times X, \quad \partial_u \mathcal{H}(u)[v] = (J\tilde{g}(u), v)_Z.$$

A smooth map φ is said to be symplectic if for all $(u, v, \tilde{v}) \in K \times X^2$, we have

$$(J\partial_u \varphi(u)[v], \partial_u \varphi(u)[\tilde{v}])_Z = (Jv, \tilde{v})_Z.$$

Remark As in the finite dimensional case, the flow map of an Hamiltonian equation is a symplectic map.

Proposition 2.3 *Whenever $s \geq 2$, the following statements hold true.*

1. *The function $h : u \in K \rightarrow Z \ni f(|u|_2^2)J^{-1}u$ is Hamiltonian with associated Hamiltonian*

$$H(u) = \int_{\mathbb{T}_{2\pi}^D} F(|u(x)|_2^2) dx,$$

where $F(u) = \int_0^u f(v) dv$.

2. *For all $j \in \llbracket 1, d \rrbracket$, the linear operator $A_j \in \mathcal{L}(X, Z)$ of (2.2) is Hamiltonian with associated Hamiltonian*

$$I_j(u) = \frac{1}{2}(JA_j u, u)_Z.$$

3. *For all $j \in \llbracket 1, d \rrbracket$ and for all $\theta_j \in [0, 1]$, the linear operator $e^{\theta_j A_j}$ is a symplectic map.*

Proof. We have for $(u, v) \in K \times X$,

$$\partial_u H(u)[v] = (f(|u|_2^2)u, v)_Z = (Jf(|u|_2^2)J^{-1}u, v)_Z = (Jh(u), v)_Z.$$

Similarly, we have

$$\partial_u I_j(u)[v] = \frac{1}{2}(JA_j v, u)_Z + \frac{1}{2}(JA_j u, v)_Z.$$

The operator Δ_j is symmetric w.r.t. $(\cdot, \cdot)$, so the linear operator $JA_j = \begin{pmatrix} \Delta_j & 0 \\ 0 & \Delta_j \end{pmatrix}$ is symmetric w.r.t. $(\cdot, \cdot)_Z$ and $\partial_u I_j(u)[v] = (JA_j u, v)_Z$. Finally, for $v_1, v_2 \in X$, we have for $\theta_j \in \mathbb{R}$,

$$\partial_{\theta_j}(Je^{\theta_j A_j} v_1, e^{\theta_j A_j} v_2)_Z = 0$$

since the operator A_j is skew-symmetric w.r.t. $(\cdot, \cdot)_Z$. Then we conclude

$$(Je^{\theta_j A_j} v_1, e^{\theta_j A_j} v_2)_Z = (Jv_1, v_2)_Z.$$

■

2.3 Projection in a finite dimensional space

Every function p in $L^2_{\mathbb{R}}(\mathbb{T}^d)$ can be developed in Fourier series. For $k \in \mathbb{Z}^d$, we define the k^{th} Fourier coefficients of p as the function $\hat{p}_k \in \mathbb{C}$, by

$$\hat{p}_k = \int_{\mathbb{T}^d} p(x) e^{-2i\pi(k \cdot x)} dx.$$

Whenever the function p is in $H^s_{\mathbb{R}}(\mathbb{T}^d)$, we have

$$\|p\|_s^2 = \sum_{k \in \mathbb{Z}^d} (1 + |k|^2)^s |\hat{p}_k|^2$$

with $|k| = \sum_{j=1}^d |k_j|$. We directly have

$$\forall k \in \mathbb{Z}^d, \quad |\hat{p}_k|^2 \leq \frac{1}{(1 + |k|^2)^s} \sum_{k' \in \mathbb{Z}^d} (1 + |k'|^2)^s |\hat{p}_{k'}|^2 \leq \frac{\|p\|_s^2}{(1 + |k|^2)^s}.$$

We now decompose the function $u = \begin{pmatrix} p \\ q \end{pmatrix} \in X$ in Fourier series. We define the Fourier coefficients of u as

$$\forall k \in \mathbb{Z}^d, \quad \hat{u}_k = \begin{pmatrix} \hat{p}_k \\ \hat{q}_k \end{pmatrix}$$

where the $\hat{p}_k$'s and $\hat{q}_k$'s are respectively the Fourier coefficients of p and q . We recall the notation $|u|_2 = (|p|^2 + |q|^2)^{1/2}$. Then, it follows that

$$\|u\|_X^2 = \sum_{k \in \mathbb{Z}^d} (1 + |k|^2)^s |\hat{q}_k + i\hat{p}_k|^2 = \sum_{k \in \mathbb{Z}^d} (1 + |k|^2)^s |\hat{u}_k|_2^2$$

and we have

$$\forall k \in \mathbb{Z}^d, \quad |\hat{u}_k|_2^2 \leq \frac{\|u\|_X^2}{(1 + |k|^2)^s}. \quad (2.5)$$

For $L \in \mathbb{N}^*$, we introduce the $\mathbb{R}^2$ -vector space

$$Z_L = \text{Span} \left\{ x \rightarrow e^{2i\pi(k \cdot x)}, \quad k \in \mathbb{Z}^d \text{ s.t. } |k| \leq L \right\} \subset Z.$$

We denote X_L the finite dimensional vector space $X \cap Z_L$ and we introduce the projection onto X_L

$$\Pi_L : u \in X \rightarrow X_L \ni (t, x) \rightarrow \sum_{|k| \leq L} \hat{u}_k e^{2i\pi(k \cdot x)},$$

where the $\hat{u}_k$'s are the Fourier coefficients of u .

Lemma 2.4 *Let $s > d/2$. For any $u \in X$ and $L \in \mathbb{N}^*$, the following bound holds.*

$$\|u - \Pi_L u\|_Z \leq \frac{\|u\|_X}{L^s}. \quad (2.6)$$

Proof. It is clear that the Fourier coefficients of the function $u - \Pi_L u$ are $\hat{u}_k$ if $|k| > L$ and vanish if $|k| \leq L$. Thus we have for all $s > d/2$,

$$\|u - \Pi_L u\|_Z^2 = \sum_{|k| > L} |\hat{u}_k|_2^2 \leq \sum_{|k| > L} \frac{(1 + |k|^2)^s}{(1 + L^2)^s} |\hat{u}_k|_2^2 \leq \frac{\|u\|_X^2}{L^{2s}}.$$

For the sake of simplicity, we denote for all $t \in [0, T/\varepsilon]$, $u_L^\varepsilon(t) = \Pi_L u^\varepsilon(t)$, where the function u^ε is defined in (2.1). Applying Π_L on both sides of (2.2), u_L^ε satisfies the equation

$$\partial_t u_L^\varepsilon(t, x) = (\omega \cdot A) u_L^\varepsilon(t, x) + \varepsilon \Pi_L(h(u^\varepsilon(t, x))), \quad u_L^\varepsilon(0, \cdot) = \Pi_L u_0 \quad (2.7)$$

or the equation

$$\partial_t u_L^\varepsilon(t, x) = (\omega \cdot A) u_L^\varepsilon(t, x) + \varepsilon \Pi_L h(u_L^\varepsilon(t, x)) + \varepsilon \Lambda_L(t, x) \quad (2.8)$$

with initial conditions $u_L^\varepsilon(0, \cdot) = \Pi_L u_0 \in X_L$ and where we have denoted, for the sake of brevity,

$$\Lambda_L(t, x) = \Pi_L h(u^\varepsilon(t, x)) - \Pi_L h(u_L^\varepsilon(t, x)). \quad (2.9)$$

It is clear that we have

$$\forall t \in [0, T/\varepsilon], \quad \|u^\varepsilon(t) - u_L^\varepsilon(t)\|_Z \leq \frac{2\|u_0\|_X}{L^s}. \quad (2.10)$$

The next step is to bound the function Λ_L in Z and to solve a truncated version of equation (2.8) (in the next section).

Lemma 2.5 *Let $s > d/2$. We have for all $t \in [0, T/\varepsilon]$ the following bound:*

$$\|\Lambda_L(t)\|_Z \leq \frac{2C_h \|u_0\|_X}{L^s}. \quad (2.11)$$

Proof. We have for $t \in [0, T/\varepsilon]$ and $x \in \mathbb{T}^d$ the relation,

$$\Lambda_L(t, x) = \Pi_L \left(h(u^\varepsilon(t, x)) - h(u_L^\varepsilon(t, x)) \right).$$

We directly have by Proposition 2.2 (since it is clear that $u_L^\varepsilon \in K$),

$$\|\Lambda_L(t)\|_Z \leq \|h(u^\varepsilon(t)) - h(u_L^\varepsilon(t))\|_Z \leq C_h \|u^\varepsilon(t) - u_L^\varepsilon(t)\|_Z,$$

which leads to the result. ■

3 Application of the quasi-stroboscopic averaging procedure in X_L

In this part, we take $L > 0$ and we apply the quasi-stroboscopic procedure to the following equation,

$$\partial_t \tilde{u}^\varepsilon(t, x) = (\omega \cdot A) \tilde{u}^\varepsilon(t, x) + \varepsilon \Pi_L h(\tilde{u}^\varepsilon(t, x)), \quad \tilde{u}^\varepsilon(0, \cdot) = \Pi_L u_0 \in X_L \quad (3.1)$$

which is a truncated version of equation (2.8). We note that the space X_L is a separable Hilbert space equipped with the scalar product (and the inner norm)

$$\forall u, \tilde{u} \in X_L, \quad (u, \tilde{u})_{X_L} = \sum_{|k| \leq L} (1 + |k|^2)^s u_k \cdot \tilde{u}_k,$$

where we write respectively for all $x \in \mathbb{T}^d$,

$$u(x) = \sum_{|k| \leq L} u_k e^{2i\pi(k \cdot x)} \quad \text{and} \quad \tilde{u}(x) = \sum_{|k| \leq L} \tilde{u}_k e^{2i\pi(k \cdot x)}.$$

More precisely, with this norm, the space X_L is a Banach Algebra.

As in Chapter 3, we denote by $X_L^\mathbb{C}$ the complexification of the space X_L equipped with the norm (again denoted) $\|\cdot\|_{X_L}$, by K_L the open space

$$K_L = \{u \in X_L, \text{ s.t. } \|u\|_{X_L} \leq M\}$$

where $M := 2 \|\Pi_L u_0\|_{X_L}$. For $\rho > 0$, we denote by $K_{L,\rho}$ the open enlargement of K_L of size ρ :

$$K_{L,\rho} = \{u_1 + u_2, (u_1, u_2) \in K_L \times X_L^\mathbb{C}, \text{ s.t. } \|u_2\|_{X_L} \leq \rho\}.$$

We denote by $\mathbf{A}_\rho$ the set of bounded analytic functions on $K_{L,\rho}$ (see definition 2.1 of Chapter 3) equipped with the norm $\|\cdot\|_{\mathbf{A}_\rho}$. Finally, for $s > d/2$, we introduce the notation

$$\xi_s = \sum_{k \in \mathbb{Z}^D} (1 + |k|^2)^{-s} < \infty.$$

Proposition 3.1 *We assume that the radius of analitycity of the function f , R_0 , satisfies $R_0 > M^2 \xi_s$. Then, the function h defined in equation (2.2) is analytic in the sense of definition 2.1 of Chapter 3. More precisely, we have*

$$\|\Pi_L h\|_{\mathbf{A}_{4R}} \leq \tilde{C}$$

with $R = \frac{1}{4} \left(\sqrt{\frac{R_0}{\xi_s}} - M \right)$ and $\tilde{C} = (4R + M) \|f\|_\infty$ two positives constants independent of L .

Proof. Let $\tilde{u} = \begin{pmatrix} p \\ q \end{pmatrix} \in K_{L,4R}$, such that for all $x \in \mathbb{T}^d$, $p(x) = \sum_{|k| \leq L} p_k e^{2i\pi k \cdot x}$ and $q(x) = \sum_{|k| \leq L} q_k e^{2i\pi k \cdot x}$. Then, we clearly have

$$\sup_{x \in \mathbb{T}^d} |\tilde{u}(x)|_2^2 = \sup_{x \in \mathbb{T}^d} (|p(x)|^2 + |q(x)|^2) \leq \left(\sum_{|k| \leq L} |p_k| \right)^2 + \left(\sum_{|k| \leq L} |q_k| \right)^2$$

Using Cauchy-Schwarz inequality, it follows

$$\left(\sum_{|k| \leq L} |p_k| \right)^2 \leq \sum_{|k| \leq L} (1 + |k|^2)^{-s} \sum_{|k| \leq L} (1 + |k|^2)^s |p_k|^2 \leq \xi_s \sum_{|k| \leq L} (1 + |k|^2)^s |p_k|^2$$

and we have the same estimates for $\left(\sum_{|k| \leq L} |q_k| \right)^2$ which leads to

$$\sup_{x \in \mathbb{T}^d} |\tilde{u}(x)|_2^2 \leq \xi_s \|\tilde{u}\|_{X_L}^2 \leq \xi_s (M + 4R)^2 = R_0$$

so $|\tilde{u}|_2^2$ is in the analytic domain of f . Hence we can write, for any function $\tilde{v} \in X_L^{\mathbb{C}}$ such that $\|\tilde{v}\|_{X_L}$ is small enough, the relation

$$\Pi_L (h(\tilde{u} + \tilde{v}) - h(\tilde{u}) - \partial_u h(\tilde{u})\tilde{v}) = \Pi_L \int_0^1 (1-t) \partial_u^2 h(\tilde{u} + t\tilde{v})[\tilde{v}, \tilde{v}].$$

One can say that the right hand side is an $o(\|\tilde{v}\|_{X_L})$. Hence, the function $\tilde{u} \rightarrow \Pi_L h(\tilde{u})$ is analytic in the sense of Definition 2.1 from Chapter 3. Moreover, for all $\tilde{u} \in K_{L,4R}$ the function $\Pi_L h(\tilde{u})$ is bounded as

$$\|\Pi_L h\|_{\mathbf{A}_{4R}} = \sup_{\tilde{u} \in K_{L,4R}} \|\Pi_L h(\tilde{u})\|_{X_L} \leq C_h \sup_{\tilde{u} \in K_{L,4R}} \|\tilde{u}\|_{X_L},$$

which ends the proof. ■

3.1 A theorem of quasi-stroboscopic averaging

In order to apply Theorem 2.13 of Chapter 3, we transform equation (3.1) with the change of unknown $v(t, x) = e^{t\omega \cdot A} \tilde{u}(t, x)$ for $t \in [0, T/\varepsilon]$ and $x \in \mathbb{T}^d$. We obtain the equivalent equation

$$\frac{d}{dt} v(t, x) = \varepsilon g(t, v(t, x)) := \varepsilon e^{t\omega \cdot A} \Pi_L h(e^{-t\omega \cdot A} v(t, x)) \quad (3.2)$$

with the initial conditions

$$v(0, \cdot) = \Pi_L u_0 \in X_L$$

The function g defined above is clearly quasi-periodic since the operator $e^{\theta_j A_j}$ is 1-periodic with respect to θ_j . In accordance with the previous chapter, we shall denote (for $\rho > 0$ and $\sigma > 0$), $\mathcal{A}_\rho^\omega$ the set of functions in $\mathcal{C}^{d+1}(\mathbb{T}^d, \mathbf{A}_\rho)$ quasi-periodic, D_σ the enlargement of the d -dimensional torus $\mathbb{T}^d$

$$D_\sigma = \left\{ \theta = \theta_1 + i\theta_2, (\theta_1, \theta_2) \in \mathbb{T}^d \times \mathbb{R}^d, \|\theta_2\|_\infty < \sigma \right\}, \quad (3.3)$$

and $\mathcal{A}_{\sigma,\rho}^\omega$ the set of function in $\mathcal{A}_\rho^\omega$ which are analytic on D_σ and for which the associated function is bounded. The set $\mathcal{A}_{\sigma,\rho}^\omega$ is then a Banach space with the norm $\|\cdot\|_{\sigma,\rho}$ defined in Chapter 3 (section 2.3).

Remark 3.2 *To every quasi-periodic functions ϕ we associate a function $\Phi \in \mathcal{C}^{d+1}(\mathbb{T}^d, \mathbf{A}_\rho)$ such that $\phi(t) = \Phi(t\omega)$. It is this associated function Φ which has to be bounded in the definition of $\mathcal{A}_{\sigma,\rho}^\omega$.*

Remark 3.3 *In this setting, each operator A_j is bounded. It is actually essential to apply Theorem 2.13 of Chapter 3. Indeed, the operator $e^{\theta_j A_j}$ is then analytic with respect to θ_j .*

Lemma 3.4 *The operator $\omega \cdot A \in \mathcal{L}(X_L)$ is upper bounded by*

$$\|\omega \cdot A\|_{\mathcal{L}(X_L)} \leq |\omega|_2 4\pi^2 L^2$$

with $|\omega|_2 = \left(\sum_{j=1}^d |\omega_j|^2 \right)^{1/2}$.

Proof. Let $u \in X_L$ such that for all $x \in \mathbb{T}^d$,

$$u(x) = \sum_{|k| \leq L} u_k e^{2i\pi(k \cdot x)} := \sum_{|k| \leq L} \begin{pmatrix} p_k \\ q_k \end{pmatrix} e^{2i\pi(k \cdot x)}.$$

We then define by p (respectively $q, \tilde{p}, \tilde{q}$) the function with the coefficients p_k (respectively $q_k, \tilde{p}_k, \tilde{q}_k$), namely

$$\forall x \in \mathbb{T}^d, \quad p(x) = \sum_{|k| \leq L} p_k e^{2i\pi(k \cdot x)}.$$

We then have

$$\forall x \in \mathbb{T}^d, \quad \Delta_j p(x) = -4\pi^2 \sum_{|k| \leq L} (k_j)^2 p_k e^{2i\pi(k \cdot x)}.$$

Using the definition for the scalar product in X_L , we obtain

$$\begin{aligned} \|\omega_j A_j u\|_{X_L}^2 &= |\omega_j|^2 (4\pi^2)^2 \sum_{|k| \leq L} (1 + |k|^2)^s \left((k_j)^4 p_k^2 + (k_j)^4 q_k^2 \right) \\ &\leq |\omega_j|^2 (4\pi^2)^2 L^4 \|u\|_{X_L}^2, \end{aligned}$$

which gives the result. ■

By the Stone's theorem, we directly have the relation

$$\sup_{\theta \in \mathbb{T}^d} \|e^{\theta \cdot A}\|_{\mathcal{L}(X_L)} \leq 1.$$

Besides, whenever θ is in D_σ for $\sigma > 0$, the following corollary holds.

Corollary 3.5 *For all $\sigma > 0$ we have,*

$$\sup_{\theta \in D_\sigma} \|e^{\theta \cdot A}\|_{\mathcal{L}(X_L)} \leq e^{\sigma d 4\pi^2 L^2} \quad \text{and} \quad \sup_{\theta \in D_\sigma} \|e^{\theta \cdot A} - e^{\Re(\theta) \cdot A}\|_{\mathcal{L}(X_L)} \leq e^{\sigma d 4\pi^2 L^2} - 1,$$

Proof. This is a direct consequence of Lemma 3.4 and Definition (3.3) for D_σ . ■

We are now ready to assert that the function g defined in equation (3.2) is a bounded analytic and quasi-periodic function of $\mathbf{A}_{2\mathbf{R}}$.

Proposition 3.6 *We assume that the function f of equation (1.2) is real-analytic with R_0 , its radius of analyticity satisfying $R_0 > M^2 \xi_s$ and that the vector of frequencies ω satisfies the strong non-resonance condition 1.1. Then, introducing*

$$W = \frac{1}{8d\pi^2 L^2} \ln \left(\frac{4R + M}{2R + M} \right) \tag{3.4}$$

the function g defined in equation (3.2) is in $\mathcal{A}_{2W, 2R}$ and

$$\|g\|_{2W, 2R} \leq C := \frac{4R + M}{2R + M} \tilde{C}, \tag{3.5}$$

where $\tilde{C}$ is defined in Proposition 3.1.

Proof. First of all, the analyticity on f gives, by Proposition 3.1, that $\Pi_L h$ is bounded in $\mathbf{A}_{4R}$ by $\tilde{C}$. Let $v = v_1 + v_2 \in K_{L, 2R}$ such that $v_1 \in K_L$, $\|v_2\|_{X_L} \leq 2R$ and $\theta \in D_{2W}$. Then, $e^{\Re(\theta) \cdot A} v_1$ belongs to K_L and using Corollary 3.5, we compute

$$\begin{aligned} \|e^{\theta \cdot A} v - e^{\Re(\theta) \cdot A} v_1\|_{X_L} &\leq \|e^{\theta \cdot A} - e^{\Re(\theta) \cdot A}\|_{\mathcal{L}(X_L)} \|v_1\|_{X_L} + \|e^{\theta \cdot A}\|_{\mathcal{L}(X_L)} \|v_2\|_{X_L} \\ &\leq (e^{8d\pi^2 L^2 W} - 1)M + e^{8d\pi^2 L^2 W} 2R \\ &\leq e^{8d\pi^2 L^2 W} (2R + M) - M \leq 4R. \end{aligned}$$

Finally g is in $\mathcal{A}_{2W, 2R}^\omega$ and is bounded by the following:

$$\|g\|_{2W, 2R} \leq e^{8d\pi^2 L^2 W} \|h\|_{4R} = C. \tag{3.6}$$

We are now in position to apply the quasi-stroboscopic procedure developed in Chapter 3 to equation (3.2). ■

Theorem 3.7 *We assume that the function f of equation (1.2) is real-analytic with $R_0 > M^2 \xi_s$, its radius of analyticity and that the vector of frequencies ω satisfies the strong non-resonance condition 1.1. Then, there exists $\tilde{\varepsilon}(L) > 0$ under the form*

$$\tilde{\varepsilon}(L) = \min \left(\varepsilon^*, \varepsilon_1, \beta_1 L^{-2(\nu+d)} \right) \quad (3.7)$$

such that for all $0 < \varepsilon < \tilde{\varepsilon}(L)$, the solution $v^\varepsilon \in \mathcal{C}^0([0, T/\varepsilon], X_L)$ of (3.2) remains in K_L and there is a quasi-stroboscopic change of variable ϕ^ε and a vector field $\mathcal{G}^\varepsilon$ such that for all $t \in [0, T/\varepsilon]$,

$$\|v^\varepsilon(t) - \phi^\varepsilon(t, \psi^\varepsilon(t, \Pi_L u_0))\|_{X_L} \leq C_2 L^{-2(\nu+d)} \exp \left(-\beta(L^{2(\nu+d)} \varepsilon)^{-\eta} \right) \quad (3.8)$$

where $\eta = \frac{1}{\nu+d+1}$, ε^ is the constant from Proposition 1.3, ψ^ε is the solution of the autonomous equation*

$$\frac{d}{dt} \psi^\varepsilon(t, \Pi_L u_0) = \varepsilon \mathcal{G}^\varepsilon(\psi^\varepsilon(t, \Pi_L u_0))$$

and where C_2, β, β_1 and ε_1 are positive constants independent of L and ε .

Proof. The proof is an application of Theorem 3.9 from Chapter 3, where we carefully write the constants. By assumption, the vector of frequencies ω satisfies the condition of strong non-resonance and the function g is bounded by C in $\mathcal{A}_{2W, 2R}$. Hence, whenever $0 < \varepsilon < \tilde{\varepsilon}(L)$, there is $\tilde{C}_1$ and $\tilde{C}_2$ such that

$$\forall t \in [0, T/\varepsilon], \quad \|v^\varepsilon(t) - \phi^\varepsilon(t, \psi^\varepsilon(t, \Pi_L u_0))\|_{X_L} \leq \tilde{C}_1 \exp \left(-\tilde{C}_2 \varepsilon^{-\eta} \right)$$

where

- From Theorem 3.7 of Chapter 3 :

$$\begin{aligned} \tilde{\varepsilon}(L) &= \min \left(\varepsilon^*, (e\kappa)^{-1}, \left(\frac{R}{2TC} \right)^{\frac{1}{\eta}} (e\kappa)^{-1}, \frac{R}{8Ce^2 m! 2^m} \right), \\ m &= \lfloor \nu + d + 1 \rfloor, \quad \tilde{C}_1 = 4Ce\kappa^{-1}, \quad \tilde{C}_2 = \frac{1}{2}(e\kappa)^{-\eta}, \\ \kappa &= 16 C \kappa_0 R^{-1} W^{-(\nu+d)}, \end{aligned}$$

- $\kappa_0 = \frac{\nu! 4^d}{\gamma \pi^{\nu+d+1}}$ (Lemma 2.10 of Chapter 3)
- $W = \frac{1}{8d\pi^2 L^2} \ln \left(\frac{4R+M}{2R+M} \right)$, $C = \frac{4R+M}{2R+M} \tilde{C}$ (Relations (3.4) and (3.5)),
- $R = \frac{1}{4} \left(\sqrt{\frac{R_0}{\xi_s}} - M \right)$, $\tilde{C} = (4R+M) \|f\|_\infty$ (Proposition 3.1).

We recall that ε^* is defined in Proposition 1.3. The proof is then straightforward noticing that L is only present in W hence κ . ■

3. APPLICATION OF THE QUASI-STROBOSCOPIC AVERAGING PROCEDURE IN X_L

Remark 3.8 *The quasi-stroboscopic procedure is then deduced for Equation (2.2). Indeed, taking $\Phi^\varepsilon(t) = e^{-t\omega \cdot A} \phi^\varepsilon(t)$ which is a quasi-periodic function, we obtain*

$$\|\tilde{u}^\varepsilon(t) - \Phi^\varepsilon(t, \psi^\varepsilon(t, \Pi_L u_0))\|_{X_L} \leq C_2 L^{-2(\nu+d)} \exp\left(-\beta(L^{2(\nu+d)} \varepsilon)^{-\eta}\right) \quad (3.9)$$

We now show that the function $\tilde{u}^\varepsilon$ solution of equation (3.1) is close to u_L^ε solution of equation (2.8).

Proposition 3.9 *Let $s > d/2$. The solutions $\tilde{u}^\varepsilon$ and u_L^ε of equations (3.1) and (2.8) satisfy*

$$\forall t \in [0, T/\varepsilon], \quad \|\tilde{u}^\varepsilon(t) - u_L^\varepsilon(t)\|_Z \leq \frac{2C_h \|u_0\|_X e^{C_h T}}{L^s} \varepsilon. \quad (3.10)$$

Proof. For $t \in [0, T/\varepsilon]$, we introduce $v_L^\varepsilon(t) = e^{-t\omega \cdot A} u_L^\varepsilon(t)$. Hence, the functions $\tilde{v}^\varepsilon$ and v_L^ε satisfies respectively equations (3.2) and

$$\partial_t v_L^\varepsilon(t) = \varepsilon e^{-t\omega \cdot A} (\Pi_L h(e^{t\omega \cdot A} v_L^\varepsilon(t))) + \varepsilon e^{-t\omega \cdot B} \Lambda_L(t, x).$$

Since $\|e^{-t\omega \cdot A}\|_{\mathcal{L}(Z)} \leq 1$, using relation (2.11) (in Lemma 2.5), it is clear that for all $t \in [0, T/\varepsilon]$,

$$\|\tilde{v}^\varepsilon(t) - v_L^\varepsilon(t)\|_Z \leq \varepsilon \int_0^t C_h \|\tilde{v}^\varepsilon(\tau) - v_L^\varepsilon(\tau)\|_Z d\tau + \varepsilon \frac{2C_h \|u_0\|_X}{L^s}$$

A Grönwall Lemma ends the proof. ■

Finally, combining the relations (2.10), (3.9) and (3.10) we obtain

1. A quasi-periodic change of variable Φ^ε ,
2. A vector field $\mathcal{G}^\varepsilon$ with ψ^ε the solution of the autonomous equation

$$\frac{d}{dt} \psi^\varepsilon(t, \Pi_L u_0) = \varepsilon \mathcal{G}^\varepsilon(\psi^\varepsilon(t, \Pi_L u_0)),$$

such that for all $t \in [0, T/\varepsilon]$, $L \in \mathbb{N}^*$, we have

$$\|u^\varepsilon(t) - \Phi^\varepsilon(t, \psi^\varepsilon(\Pi_L u_0))\|_Z \leq \frac{C_1}{L^s} + \frac{C_2}{L^{2(\nu+d)}} e^{-\beta(L^{2(\nu+d)} \varepsilon)^{-\eta}}, \quad (3.11)$$

where $C_1 = 2 \|u_0\|_X (1 + C_h e^{C_h T})$ is a constant independent of L and ε (we have here considered that $\varepsilon \leq 1$).

4 Quasi-periodic averaging with polynomial remainder estimates

In this part, we would like to optimize in estimation (3.11), the choice of the integer L in order to obtain exponential remainder estimates. However, the choice of such a L imposes that for a fixed $\varepsilon > 0$, s has to be large enough (which implies that the initial condition u_0 has a high regularity). Besides, whenever ε is close to zero, s goes to infinity.

The fact that s has to be large when ε is small is an issue. However, the quasi-stroboscopic averaging remains true for any values of s with polynomial remainder estimates. From now on, the quantity s is fixed and satisfies $s \geq 4(\nu + d)$. We then define $n \in \mathbb{N}^*$ an integer such that

$$n \leq \frac{s}{4(\nu + d)}.$$

We show that we have unconditionally polynomial remainder estimates (in the sense that they are valid for all values of ε).

Lemma 4.1 *Assume that $s \geq 4(\nu + d)$ and consider $n \leq \frac{s}{4(\nu + d)}$, $L_\varepsilon = \lfloor \varepsilon^{-ns^{-1}} \rfloor + 1 \in \mathbb{N}^*$. Then estimation (3.11) reads for all $t \in [0, T/\varepsilon]$,*

$$\|u^\varepsilon(t) - \Phi^\varepsilon(t, \psi^\varepsilon(\Pi_L u_0))\|_Z \leq (C_1 + \tilde{C}_2)\varepsilon^n, \quad (4.1)$$

where C_1 and $\tilde{C}_2$ are constants independent of ε .

Proof. First of all, since $L_\varepsilon \geq \varepsilon^{-ns^{-1}}$, we clearly have

$$\frac{1}{L_\varepsilon^s} \leq \varepsilon^n.$$

Secondly, using the inequality $e^{-x} \leq \frac{m!}{x^m}$ (valid for all $x > 0$ and $m \in \mathbb{N}$) with $x = \beta(L^{2(\nu+d)}\varepsilon)^{-\eta}$, we have for all $m \in \mathbb{N}$,

$$\frac{1}{L_\varepsilon^{2(\nu+d)}} e^{-\beta(L^{2(\nu+d)}\varepsilon)^{-\eta}} \leq \frac{m!}{\beta^m} L^{2(\nu+d)(m\eta-1)} \varepsilon^{m\eta}$$

The choice of m is crucial here and we consider

$$m := \left\lfloor \frac{1}{\eta} \left(\frac{n - \frac{2n}{s}(\nu + d)}{1 - \frac{2n}{s}(\nu + d)} \right) \right\rfloor + 1 \in \mathbb{N}^*.$$

This quantity is well defined since $(2(\nu + d)s^{-1} \leq 1/2$ and $2n(\nu + d)s^{-1} \leq 1/2$. Hence, we have on the one hand

$$m\eta \geq \left(\frac{n - \frac{2n}{s}(\nu + d)}{1 - \frac{2n}{s}(\nu + d)} \right) \geq \left(1 + \frac{n - 1}{1 - \frac{2n}{s}(\nu + d)} \right) \geq 1$$

and on the other hand

$$m\eta \left(1 - \frac{2n}{s}(\nu + d)\right) + \frac{2n}{s}(\nu + d) = \frac{n}{s}(2(\nu + d)(1 - m\eta) + m\eta) \geq n$$

which leads to

$$\varepsilon^{\frac{2n}{s}(\nu+d)(1-m\eta)+m\eta} \leq \varepsilon^n.$$

Since $m\eta - 1 \geq 0$, we consider

$$L_\varepsilon^{2(\nu+d)(m\eta-1)} \leq (\varepsilon^{-ns^{-1}} + 1)^{2(\nu+d)(m\eta-1)} \leq (2\varepsilon^{-ns^{-1}})^{2(\nu+d)(m\eta-1)}.$$

Combining all inequalities together, we obtain

$$\frac{1}{L_\varepsilon^{2(\nu+d)}} e^{-\beta(L^{2(\nu+d)}\varepsilon)^{-\eta}} \leq \frac{m!}{\beta^m} 2^{2(\nu+d)(m\eta-1)} \varepsilon^n.$$

Then, estimate (4.1) holds with $\tilde{C}_2 = \frac{m!}{\beta^m} 2^{2(\nu+d)(m\eta-1)} C_2$. ■

We are now ready to give the main result of quasi-stroboscopic averaging for the semi-linear Schrödinger equation.

Theorem 4.2 *Let $s > \max(d/2, 4(\nu + d))$ and $n \leq \frac{s}{4(\nu+d)}$. Assume that*

$$0 < \varepsilon < \min \left(\varepsilon^*, \varepsilon_1, \frac{\beta_1^2}{2^{4(\nu+d)}} \right)$$

where ε^* is introduced in Proposition 1.3 and the constants ε_1 and β_1 are defined in Theorem 3.7. Then there exist a quasi-periodic analytic change of variable Φ^ε and a vector $\mathcal{G}^\varepsilon$ such that for all $t \in [0, T/\varepsilon]$,

$$\|u^\varepsilon(t) - \Phi^\varepsilon(t, \psi^\varepsilon(t, \Pi_L u_0))\|_Z \leq \mathbf{C} \varepsilon^n \quad (4.2)$$

where u^ε is the solution of equation (2.2), $\eta = \frac{1}{\nu+d+1}$, ψ^ε is the solution of the autonomous equation

$$\frac{d}{dt} \psi^\varepsilon(t, \Pi_L u_0) = \varepsilon \mathcal{G}^\varepsilon(\psi^\varepsilon(t, \Pi_L u_0))$$

and where $\mathbf{C}$ is a positive constants independent of ε .

Proof. We prove this theorem by applying Theorem 3.7 in a finite dimensional space X_L with L well-chosen. We define

$$L_\varepsilon = \lfloor \varepsilon^{-ns^{-1}} \rfloor + 1 \in \mathbb{N}^*$$

Then for every $\varepsilon \leq \varepsilon_2 := \frac{\beta_1^2}{2^{4(\nu+d)}}$, the following holds true.

$$L_\varepsilon \leq 2\varepsilon^{-\frac{n}{s}} \leq 2\varepsilon^{-\frac{1}{4(\nu+d)}} \leq 2\varepsilon^{\frac{1}{4(\nu+d)}} \varepsilon^{-\frac{1}{2(\nu+d)}} \leq \beta_1^{\frac{1}{2(\nu+d)}} \varepsilon^{-\frac{1}{2(\nu+d)}},$$

where we have used the relation $ns^{-1} \leq \frac{1}{4(\nu+d)}$. It is clear that this relation is equivalent to $\varepsilon \leq \beta_1 L_\varepsilon^{-2(\nu+d)}$. Hence, for $\varepsilon < \min(\varepsilon^*, \varepsilon_1, \beta_1^2(2^{-4(\nu+d)}))$ we have $\varepsilon \leq \tilde{\varepsilon}(L_\varepsilon)$ where $\tilde{\varepsilon}$ is defined by (3.7) in Theorem 3.7.

We then apply Theorem 3.7 in the space X_{L_ε} which provides Φ^ε and ψ^ε such that relation (3.11) holds for all $t \in [0, T/\varepsilon]$:

$$\|u^\varepsilon(t) - \Phi^\varepsilon(t, \psi^\varepsilon(t, \Pi_{L_\varepsilon} u_0))\|_Z \leq \frac{C_1}{L^s} + \frac{C_2}{L^{2(\nu+d)}} e^{-\beta(L^{2(\nu+d)}\varepsilon)^{-\eta}}.$$

Finally, applying Lemma 4.1, we obtain

$$\|u^\varepsilon(t) - \Phi^\varepsilon(t, \psi^\varepsilon(t, \Pi_{L_\varepsilon} u_0))\|_Z \leq \mathbf{C}\varepsilon^n$$

where $\mathbf{C} = C_1 + \tilde{C}_2$ (we recall that the constant $\tilde{C}_2$ is given in the proof of Lemma 4.1) is a constant independent of ε . ■

Bibliography

- [1] Serge Alinhac and Patrick Gérard. *Pseudo-differential Operators and the Nash-Moser Theorem*, volume 82 of *Graduate Studies in Mathematics*. American Mathematical Society, 1991.
- [2] V. I. Arnold. Proof of a theorem of A. N. Kolmogorov on the preservation of conditionally periodic motions under a small perturbation of the Hamiltonian. *Uspehi Mat. Nauk*, 18(5 (113)):13–40, 1963.
- [3] Dario Bambusi. Birkhoff normal form for some nonlinear PDEs. *Comm. Math. Phys.*, 234(2):253–285, 2003.
- [4] N. N. Bogoliubov and Y. A. Mitropolsky. *Asymptotic methods in the theory of non-linear oscillations*. Translated from the second revised Russian edition. International Monographs on Advanced Mathematics and Physics. Hindustan Publishing Corp., Delhi, Gordon and Breach Science Publishers, New York, 1961.
- [5] N. N. Bogolyubov and Yu. A. Mitropolskiĭ. *Asimptotičeskie metody v teorii nelineĭnykh kolebaniĭ*. Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow, 1955.
- [6] Jean Bourgain. Construction of approximative and almost periodic solutions of perturbed linear Schrödinger and wave equations. *Geom. Funct. Anal.*, 6(2):201–230, 1996.
- [7] J. C. Butcher. An algebraic theory of integration methods. *Math. Comp.*, 26:79–106, 1972.
- [8] J. C. Butcher. *The numerical analysis of ordinary differential equations*. A Wiley-Interscience Publication. John Wiley & Sons, Ltd., Chichester, 1987. Runge Kutta and general linear methods.
- [9] M. P. Calvo, Ph. Chartier, A. Murua, and J. M. Sanz-Serna. Numerical stroboscopic averaging for ODEs and DAEs. *Appl. Numer. Math.*, 61(10):1077–1095, 2011.

- [10] Mari Paz Calvo, Philippe Chartier, Ander Murua, and Jesús María Sanz-Serna. A stroboscopic numerical method for highly oscillatory problems. In Björn Engquist, Olof Runborg, and Yen-Hsi R. Tsai, editors, *Numerical Analysis of Multiscale Computations*, volume 82 of *Lecture Notes in Computational Science and Engineering*, pages 71–85. Springer Berlin Heidelberg, 2012.
- [11] M.P. Calvo, Ph. Chartier, A. Murua, and J.M. Sanz-Serna. Numerical stroboscopic averaging for odes and daes. *Applied Numerical Mathematics*, 61(10):1077 – 1095, 2011.
- [12] Rémi Carles and Erwan Faou. Energy cascades for NLS on the torus. *Discrete Contin. Dyn. Syst.*, 32(6):2063–2077, 2012.
- [13] F. Castella, Ph. Chartier, F. Méhats, and A. Murua. Stroboscopic averaging for the nonlinear Schrödinger equation. *Found. Comput. Math.*, 15(2):519–559, 2015.
- [14] Arthur Cayley. On the theory of the analytic forms called trees. *Philos. Mag.*, XIII:172–176, 1857.
- [15] T. Cazenave. *Semilinear Schrödinger Equations*. Courant lecture notes in mathematics. American Mathematical Society, 2003.
- [16] P. Chartier, A. Murua, and J. M. Sanz-Serna. Higher-order averaging, formal series and numerical integration I: B-series. *Found. Comput. Math.*, 10(6):695–727, 2010.
- [17] P. Chartier, A. Murua, and J. M. Sanz-Serna. Higher-order averaging, formal series and numerical integration II: The quasi-periodic case. *Found. Comput. Math.*, 12(4):471–508, 2012.
- [18] P. Chartier, A. Murua, and J. M. Sanz-Serna. Higher-order averaging, formal series and numerical integration III: error bounds. *Found. Comput. Math.*, 15(2):591–612, 2015.
- [19] Philippe Chartier, N. Mauser, Florian Méhats, and Y. Zhang. Solving highly-oscillatory nls with sam: numerical efficiency and geometric properties. Accepted for publication in DCDS.
- [20] Philippe Chartier, Ander Murua, and Jesús María Sanz-Serna. A formal series approach to averaging: exponentially small error estimates. *Discrete Contin. Dyn. Syst.*, 32(9):3009–3027, 2012.
- [21] Alexis Clairaut. Mémoire sur l’orbite apparent du soleil autour de la terre an ayant égard aux perturbations produites par les actions de la lune et des planètes principales. *Mémoires de l’Académie des Sciences (Paris)*, pages 521–564, 1754.

-
- [22] David Cohen, Ernst Hairer, and Christian Lubich. Modulated Fourier expansions of highly oscillatory differential equations. *Found. Comput. Math.*, 3(4):327–345, 2003.
 - [23] David Cohen, Ernst Hairer, and Christian Lubich. Long-time analysis of nonlinearly perturbed wave equations via modulated Fourier expansions. *Arch. Ration. Mech. Anal.*, 187(2):341–368, 2008.
 - [24] J.D. Cole and J.K. Kevorkian. *Multiple Scale and Singular Perturbation Methods*. Applied Mathematical Sciences Series. Springer London, Limited, 1996.
 - [25] Robert Dautray and Jacques Louis Lions. *Mathematical Analysis and Numerical Methods for Science and Technology Mathematical Analysis and Numerical Methods for Science and Technology*, volume 5. Springer, 1992.
 - [26] Pierre-Simon de Laplace. *Traité de mécanique céleste*, volume 1-5. Duprat, Courcier-Bachelier, Paris, 1779.
 - [27] James A. Ellison, Albert W. Sáenz, and H. Scott Dumas. Improved nth order averaging theory for periodic systems. *J. Differential Equations*, 84(2):383–403, 1990.
 - [28] Pierre Fatou. Sur le mouvement d’un système soumis à des forces à courtes périodes. *Bulletin de la Societe Mathematique*, 56:98–139, 1928.
 - [29] Enrico Fermi, J Pasta, and S Ulam. Studies of nonlinear problems. *Los Alamos Report LA-1940*, page 978, 1955.
 - [30] Benoît Grébert and Carlos Villegas-Blas. On the energy exchange between resonant modes in nonlinear Schrödinger equations. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 28(1):127–134, 2011.
 - [31] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving ordinary differential equations. I*, volume 8 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 1993. Nonstiff problems.
 - [32] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric numerical integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2010. Structure-preserving algorithms for ordinary differential equations, Reprint of the second (2006) edition.
 - [33] Mandelstam Leonid Issaakovitch and Nicolai Dmitrievich Papalexi. über die begründung einer method für die näherungslösung von differentialgleichungen. *Zeitschrift für Experimentelle und Theoretische Physik*, 4:117, 1934.

- [34] A. N. Kolmogorov. On conservation of conditionally periodic motions for a small change in Hamilton's function. *Dokl. Akad. Nauk SSSR (N.S.)*, 98:527–530, 1954.
- [35] N. Kryloff and N. Bogoliuboff. *Introduction to Non-Linear Mechanics*. Annals of Mathematics Studies, no. 11. Princeton University Press, Princeton, N. J., 1943.
- [36] Joseph-Louis Lagrange. *Mécanique Analytique*. Edition Albert Blanchard, Paris, 1788.
- [37] G. Leboucher. Stroboscopic averaging of highly oscillatory nonlinear wave equations. *Math. Methods Appl. Sci.*, 38(9):1746–1766, 2015.
- [38] Jia Li and Junxiang Xu. On the reducibility for a class of quasi-periodic Hamiltonian systems with small perturbation parameter. *Abstr. Appl. Anal.*, pages Art. ID 354063, 17, 2011.
- [39] Pierre Lochak and Claude Meunier. *Multiphase averaging for classical systems with applications to adiabatic theorems*, volume 72. Springer-Verlag, 1988.
- [40] R.H. Merson. An operational method for the study of integration processes. *Proc. Symp. Data Processing Weapons Research Establishment*, pages 110–111–25, 1957.
- [41] J. Moser. On invariant curves of area-preserving mappings of an annulus. *Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II*, 1962:1–20, 1962.
- [42] J. Moser. On the theory of quasiperiodic motions. *SIAM Rev.*, 8:145–172, 1966.
- [43] Amnon Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*, volume 44 of *Applied Mathematical Sciences*. Springer, 1983.
- [44] L.M. Perko. Higher order averaging and related methods for perturbed periodic and quasi-periodic systems. *SIAM Journal on Applied Mathematics*, 17(4):698–724, Jul. 1969.
- [45] Jürgen Pöschel and Eugène Trubowitz. *Inverse spectral theory*, volume 130. Harcourt Brace Jovanovich, Publishers, c1987 edition, 1987.
- [46] Michael Reed and Barry Simon. *Methods of modern mathematical physics. I*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, second edition, 1980. Functional analysis.

- [47] J. A. Sanders, F. Verhulst, and J. Murdock. *Averaging methods in nonlinear dynamical systems*, volume 59 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2007.
- [48] J. M. Sanz-Serna. Modulated Fourier expansions and heterogeneous multiscale methods. *IMA J. Numer. Anal.*, 29(3):595–605, 2009.
- [49] Terence Tao. *Nonlinear Dispersive Equations: Local and Global Analysis*. Number 106 in Conference Board of the Mathematical Science. American Mathematical Society and National Science Foundation, 2005.

