# EXPONENTIAL AVERAGING
# FOR HAMILTONIAN EVOLUTION EQUATIONS

KARSTEN MATTHIES AND ARND SCHEEL

ABSTRACT. We derive estimates on the magnitude of non-adiabatic interaction between a Hamiltonian partial differential equation and a high-frequency nonlinear oscillator. Assuming spatial analyticity of the initial conditions, we show that the dynamics can be transformed to the uncoupled dynamics of an infinite-dimensional Hamiltonian system and an anharmonic oscillator, up to coupling terms which are exponentially small in a certain power of the frequency of the oscillator. The result is derived from an abstract averaging theorem for infinite-dimensional analytic evolution equations in Gevrey spaces. Refining upon a similar result by Neishtadt for analytic ordinary differential equations, the temporal estimate crucially depends on the spatial regularity of the initial condition. The result shows to what extent the strong resonances between rapid forcing and highly oscillatory spatial modes can be suppressed by the choice of sufficiently smooth initial data. An application is provided by a system of nonlinear Schrödinger equations, coupled to a rapidly forcing single mode, representing small-scale oscillations. We provide an example showing that the estimates for partial differential equations we derive here are necessarily different from those in the context of ordinary differential equations.

## 1. INTRODUCTION

Resonances and interaction between nonlinear coupled oscillators have been a major driving force in the development of modern dynamical systems theory. Basically, at least two lines of investigation can be traced back since Poincaré's work on celestial mechanics. The first strategy emphasizes the role of interaction and tries to describe its effects. This leads to the development of normal form theory and the study of homoclinic bifurcations. On the other side, weak resonances can often be estimated to have very limited effects.

A most extreme case of coupling occurs already in Poincaré's original work on celestial mechanics, where a rapid oscillation with frequency $\Omega/\varepsilon$ is coupled to a slow motion. To what extent this coupling actually leads to dramatic instabilities of regular motion still remains a largely unsolved question today. A good approximation for the dynamics of a rapidly forced system can be found by averaging

the vector field over one period of the oscillation. Better approximations to the actual slow motion of the system can be obtained by higher-order averaging, expanding as a series in the period of the forcing. Even for analytic vector fields, these series expansions typically do not converge. Whereas the actual implications of the non-convergence still miss a general description, upper estimates on the effect, following the second line of investigation described above, have been derived in several contexts, first by Nekhoroshev, and, in a different spirit, by Neishtadt; see, e.g., [LoMe88]. Neishtadt found a transformation converting a rapidly forced *ordinary differential equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}y = f(y, \frac{t}{\varepsilon}), \quad f(y,\tau) = f(y,\tau+1)$$

into an autonomous equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{y} = \tilde{f}(\tilde{y}) + r(\tilde{y}, \frac{t}{\varepsilon}),$$

up to a remainder term $r$ that he showed to be exponentially small in the forcing frequency $1/\varepsilon$,

$$|r(\tilde{y},\tau)| \leq C\exp(-c/\varepsilon).$$

The central argument relies on analyticity of the vector field $f$. The transformations respect certain structures of the vector field. For example, whenever the vector field is Hamiltonian, the transformation can be chosen to be canonical, that is, to preserve the symplectic structure.

Here, our main concern is *partial differential equations* (PDEs). As a first, simplest example, we may think of the nonlinear heat equation

$$\partial_t u = \partial_{xx} u + f(u, \frac{t}{\varepsilon}),$$

with periodic boundary conditions on $[0,1]$, say. Trying to apply Neishtadt's averaging procedure, we realize that partial derivatives do not even define a continuous vector field on any reasonable Banach space, whereas Neishtadt's results require analyticity of the vector field. However, the difficulty is generated by spatially highly oscillatory Fourier modes, which diffusion is supposed to average out after a short initial transient. In [Mat01], the first author exploited this smoothing effect to prove a first averaging theorem for dissipative partial differential equations. Surprisingly enough, the estimate for the remainder seems to be worse than in the ODE context, no better than

$$|r(y,\tau)| \leq C\exp(-c/\varepsilon^{1/3}).$$

Here, only the nonlinearity $f$ has to be analytic. A lower estimate, which would show that infinite-dimensional dissipative equations are more sensitive to fast periodic forcing, as exhibited by the exponent $1/3$, is not available.

Conservative wave systems provide a variety of examples for partial differential equations with interactions between spatio-temporal patterns on many scales. The model equations are mostly *Hamiltonian partial differential equations*, simple prominent examples being the nonlinear Schrödinger equation, the Korteweg-deVries family, and semilinear wave equations. Oscillations naturally appear in Hamiltonian systems as the simplest form of stable motion. In turn, the starting point for an analysis of conservative wave systems is, typically, plane wave solutions $\exp(\mathrm{i}(kx - \omega(k)t))$. Quite often, the frequency $\omega$ tends to infinity when the spatial period $2\pi/k$ goes to zero. The interaction between long-wavelength, soliton-like

structures and small-scale plane waves then is one typical example of the coupling between slow motion and rapid oscillations in wave systems.

As a particular example, we think of a nonlinear Schrödinger equation,

$$i\partial_t u = \partial_{xx} u + u f(|u|^2),$$

describing the slow motion of solitons, coupled to a fast one-degree-of-freedom, anharmonic oscillator, representing weakly nonlinear, small-scale structures. With the idealization of representing small-scale structures by just one degree of freedom, we are able to deduce, as a typical consequence of our main theorems, that there is a change of coordinates which decouples the slow motion in the nonlinear Schrödinger equation from the rapid oscillations—up to remainder terms which are exponentially small in the forcing frequency; see Section 4.

This particular case is derived from a much more general averaging result on abstract Hamiltonian partial differential equations, coupled to an anharmonic *one-degree-of-freedom* Hamiltonian oscillator, with large minimal frequency. Theorem 2 shows that for spatially regular initial conditions, the coupling can be averaged up to a remainder which is again exponentially small in the minimal frequency of the oscillation. The theorem is a special case of an even more general result on abstract evolution equations, posed on Banach spaces, which "respect" high spatial regularity; see Theorem 1. In particular, we show that the results from [Mat01] for dissipative PDEs can be extended to certain conservative equations. Moreover, improving upon the corresponding results for dissipative systems, we give examples which show that the estimates, here, are necessarily different from Neishtadt's estimates.

Technically, Galerkin-type approximations are used to approximate the infinite-dimensional PDE by an ordinary differential equation. Normal form transformations are then performed just like in Neishtadt's original work, tracking the norms of the error terms produced by the transformation. Careful coupling between the number of Galerkin modes, the number of averaging steps, and the frequency of the rapid oscillations then allows us to derive exponential estimates on the remainder term.

The results are different from [Mat01] in that the assumption on a temporally regularizing effect of the evolution equation is replaced by assuming sufficiently regular initial conditions. In particular, spatial regularity suppresses high Fourier modes, which would lead to arbitrarily high oscillatory motion in the Hamiltonian PDE, generating strong resonances with the rapid forcing.

Admittedly, our results are only a small step towards an estimate on the interaction between small-scale and large-scale structures in conservative wave-systems, where on both scales typically infinitely many frequencies, and not only one frequency on the small scale, interact. Still, they give a first indication how Neishtadt's estimates for ordinary differential equations have to be combined with spatial regularity estimates in order to average spatio-temporal dynamics.

**Outline.** In Section 2, we introduce basic notation and function spaces and state our main results on exponential averaging. In Section 3, we prove the main theorems in the abstract setup. An example is provided in Section 4 by the nonlinear Schrödinger equation. Finally, in Section 5, we construct an example providing lower estimates.

## 2. Exponential averaging—main results

We state our main results on exponential averaging of abstract, infinite-dimensional evolution equations. In Section 2.1, we introduce notation used throughout and state our first main theorem. In Section 2.2, we specialize to a class of Hamiltonian evolution equations and state our second main theorem on exponential averaging. We conclude the section with a discussion of the results, drawbacks, and possible extensions; see Section 2.3.

2.1. **Averaging abstract evolution equation.** Our first result is concerned with an abstract evolution equation coupled to a rapidly rotating phase. Let $X$ be a real Banach space and let $A$ be a closed, densely defined operator, possibly unbounded, with domain $D(A)$. Denote by $S^1 = \mathbb{R}/\mathbb{Z}$ the circle. Consider the evolution equation on the phase space $X \times S^1$,

$$(2.1) \qquad \begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} u &= Au + F(u) + G(u, \psi; \varepsilon), \\ \frac{\mathrm{d}}{\mathrm{d}t}\psi &= \frac{\Omega(u)}{\varepsilon} + g(u, \psi; \varepsilon), \end{aligned}$$

where $u \in X$, $\psi \in S^1$, and $\varepsilon > 0$ denotes a small, real parameter. Smoothness assumptions on the nonlinear functions $F : X \to X$, $G : X \times S^1 \times \mathbb{R}$, $\Omega : X \to \mathbb{R}$, and $g : X \times S^1 \times \mathbb{R} \to \mathbb{R}$ will be made precise below.

We first need an assumption which allows us to approximate the infinite-dimensional evolution in (2.1) by ordinary differential equations.

**Hypothesis 2.1** (Galerkin approximation). We assume, that there exists a sequence of (Galerkin) projections $(P_m)_{m \in \mathbb{N}}$ which satisfy the following requirements:

(i) the sequence of projections converges strongly to the identity on $X$,
$$\lim_{m \to \infty} P_m u = u \text{ in } X \text{ for all } u \in X;$$

(ii) the projections $P_m$ commute with $A$ on its domain of definition
$$P_m Au = AP_m u \text{ for all } u \in D(A);$$

(iii) the operator $A$ is bounded on $\operatorname{Rg} P_m$,
$$|AP_m u|_X \le m|P_m u|_X \text{ for all } u \in X.$$

A typical example for the projections $P_m$ would be representatives of a spectral resolution of a selfadjoint operator; for example, when $A = \mathrm{i}\Delta$ on $L^2(\mathbb{R}^n)$, then $\operatorname{Rg}(P_m)$ could be chosen to be the space of functions whose Fourier transform has support in $[-m, m]^n$. We think of the sequence $P_m u$ as approximations to $u$, averaging *spatially small-scale structures*.

The next hypothesis ensures rapid rotation of the phase $\psi$.

**Hypothesis 2.2** (Rapid oscillation). We assume that there is $c > 0$ such that $\Omega(u) > c$ for all $u \in X$.

A further assumption ensures that the perturbations $G$ and $g$ have zero average over the circle $\psi \in S^1$.

**Hypothesis 2.3** (Zero mean). We assume that
$$\int_{S^1} G(u, \psi; \varepsilon)\mathrm{d}\psi = 0, \qquad \int_{S^1} g(u, \psi; \varepsilon)\mathrm{d}\psi = 0,$$

for all $u \in X$.

The spatial regularity, mentioned in the introduction, which compensates for the possible influence of highly oscillatory motion in the infinite-dimensional evolution equation, is encoded in spaces with Gevrey regularity. Gevrey spaces have proved to be useful in the analysis of semilinear parabolic problems; see, e.g., [FoTe89], [Pro91], [TBDHT96], [FeTi98]. Typically, Gevrey spaces are defined via the domain of definition of the unbounded operator $\exp(\sigma|A|^p)$. However, to define $|A|$ might require some additional work. To allow for a general statement of our result, we assume existence of a closed operator $\Gamma_{\sigma,p}$ which satisfies the characteristic properties of the backward semigroup $\exp(\sigma|A|^p)$.

**Hypothesis 2.4** (Exponential approximation)**.** Assume that there exists a closed, densely defined, boundedly invertible operator $\Gamma_{\sigma,p}$ with domain of definition

$$(2.2) \qquad\qquad \mathcal{G}_{\sigma,p} := D(\Gamma_{\sigma,p}) \subset D(A),$$

such that $\operatorname{Rg} P_m \subset \mathcal{G}_{\sigma,p}$, $\mathcal{G}_{\sigma,p}(\operatorname{Rg} P_m) = \operatorname{Rg} P_m$ for all $m$, and

$$\Gamma_{\sigma,p}Au = A\Gamma_{\sigma,p}u \text{ for all } u \in \operatorname{Rg} P_m.$$

We equip the Gevrey spaces $\mathcal{G}_{\sigma,p}$ with the graph norm

$$(2.3) \qquad\qquad |u|_{\mathcal{G}_{\sigma,p}} = |u|_X + |\Gamma_{\sigma,p}u|_X.$$

We assume that Gevrey-smooth functions in $G_{\sigma,p}$ are exponentially well approximated by the Galerkin projections $P_m$,

$$(2.4) \qquad\qquad |\Gamma_{\sigma,p}^{-1}(1 - P_m)| \le C_0 \exp(-c_0/m^p),$$

for $m$-independent constants $C_0(\sigma,p)$ and $c_0(\sigma,p)$.

In most applications, one will choose $\Gamma_{\sigma,p} = \exp(\sigma|A|^p)$ or $\Gamma_{\sigma,p} = \exp(\sigma(-A)^p)$. The latter choice has been useful in the case of sectorial operators $A$, in the parabolic context; see [Mat01]. The first choice will be exploited in the applications to infinite-dimensional Hamiltonian systems in Section 2.2.

Whenever the Galerkin approximations of an unbounded operator are derived from some spectral family decomposition, the estimate (2.4) in Hypothesis 2.4 is satisfied, in general.

For $p = 1$ and $A = \partial_{xx}$, $x \in S^1$, we may solve the heat equation backwards for a time interval $[-\sigma, 0]$ for initial conditions in $\mathcal{G}_{\sigma,1}$. In particular, functions in $\mathcal{G}_{\sigma,1}$ are therefore analytic.

A more precise measure of analyticity is found in the norm of a suitable extension to complex numbers. We therefore extend our function spaces to complex Banach spaces in the standard way.

For a general Banach space $Y$ we denote by

$$Y_{\mathbb{C}} = Y \times Y$$

the complexification with norm $|(u_1, u_2)|_{Y_{\mathbb{C}}}^2 = |u_1|_Y^2 + |u_2|_Y^2$. Linear operations are extended to $Y_{\mathbb{C}}$ as follows:

$$\begin{aligned}
\overline{(u_1, u_2)} &= (u_1, -u_2), \\
(a + bi)(u_1, u_2) &= (au_1 - bu_2, bu_1 + au_2), \\
L(u_1, u_2) &= (Lu_1, Lu_2) \text{ for } L \in L(Y, Y), \\
L(u_1, u_2) &= Lu_1 + iLu_2 \text{ for } L \in L(Y, \mathbb{R}).
\end{aligned}$$

Complexification of the circle is understood as the quotient of the complexification of the real line $\mathbb{C}$ by real integer translations $S^1_{\mathbb{C}} := \mathbb{C}/\mathbb{Z}$.

Analyticity properties of functions are quantified by the help of extensions to the complexified domain. We therefore introduce a short notation. For any open subset $U$ of a Banach space $Y$, we define the open complex $\delta$-extension $U + \delta$ for any $\delta > 0$ by

$$(2.5) \qquad U + \delta := \{y \in Y_{\mathbb{C}}| \inf_{u \in U} |u - y|_Y < \delta\}.$$

Smoothness properties of the nonlinearities can now be made precise.

**Hypothesis 2.5** (Analyticity of nonlinearities). There is a Gevrey class $Y = \mathcal{G}_{\sigma,p}$ with $\sigma, p > 0$ and a constant $\delta > 0$ for the size of the complex extension such that the following properties of the nonlinearities hold.

The nonlinearities are analytic and bounded on bounded subsets when considered on Gevrey spaces, extended in the complex direction,

$$F : (Y + \delta) \to Y_{\mathbb{C}}, \qquad \Omega : (Y_{\mathbb{C}} + \delta) \to \mathbb{C},$$
$$G : (Y + \delta) \times (S^1 + \delta) \times \mathbb{R} \to Y_{\mathbb{C}}, \qquad g : (Y + \delta) \times (S^1 + \delta) \times \mathbb{R} \to \mathbb{C}.$$

In addition, all of the above statements are assumed to hold when the space of Gevrey regularity $Y = \mathcal{G}_{\sigma,p}$ is replaced by $Y = X$.

The main objective, here, is to look for a coordinate transformation which would decouple fast and slow motion. A complete decoupling would result in an *adiabatic elimination* of the rapid phase. We will see in Theorem 1, below, that an almost complete decoupling is indeed possible for small values of $\varepsilon$. The remainder terms, coupling, in particular, the fast phase to the slow motion of $u$, reflect the *non-adiabatic effects*.

We are now ready to state our first main result.

**Theorem 1.** *Assume Hypothesis 2.1 on the existence of Galerkin approximations, Hypothesis 2.2 for a minimal large oscillation frequency, Hypothesis 2.3 on zero mean of the non-adiabatic terms, Hypothesis 2.4 on the exponential approximation and Hypothesis 2.5 on analyticity of the nonlinearities in some Gevrey class $\mathcal{G}_{\sigma,p} + \delta$.*

*Then, for any ball of radius $R$ in $\mathcal{G}_{\sigma,p} \times S^1$, there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$, the following assertions hold.*

*There exists a near-identity transformation $\mathrm{id} + \varepsilon W$, defined on the ball $B_R \times S^1 \subset X \times S^1$, which eliminates adiabatically the fast phase, up to an exponentially small non-adiabatic effect.*

*More precisely, the transformation is analytic on the ball $B_R$ in the complex extensions $(\mathcal{G}_{\sigma,p} \times S^1) + \delta'$ and $(X \times S^1) + \delta'$ for some $\delta' > 0$.*

*In the new variables $(\tilde{u}, \phi)$, the evolution equation (2.1) reads*

$$(2.6) \qquad \begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\tilde{u} &= A\tilde{u} + F(\tilde{u}) + \tilde{F}(\tilde{u}; \varepsilon) + \alpha(\tilde{u}, \phi; \varepsilon), \\ \frac{\mathrm{d}}{\mathrm{d}t}\phi &= \frac{1}{\varepsilon}(\Omega(\tilde{u}) + \tilde{\Omega}(\tilde{u}; \varepsilon)) + \beta(\tilde{u}, \phi; \varepsilon). \end{aligned}$$

*The transformed nonlinearities $\alpha$, $\beta$ and $\tilde{\Omega}$ are bounded on the ball $B_R$ in $X \times S^1$ and $B_R$ respectively, uniformly in $0 < \varepsilon < \varepsilon_0$. The adiabatic correction $\tilde{F}$ is even $O(\epsilon^{\frac{p}{p+1}})$ on the ball $B_R$ in $X$.*

*The non-adiabatic interaction terms $\alpha$ and $\beta$ are exponentially small in $\varepsilon$. More precisely, there exist constants $c_1, C_1 > 0$ such that*

$$(2.7) \qquad |\alpha(\tilde{u}, \phi; \varepsilon)|_X + |\beta(\tilde{u}, \phi; \varepsilon)| < C_1 \exp\left(-c_1 \varepsilon^{-\frac{p}{1+p}}\right),$$

*for all $\tilde{u} \in B_R \subset \mathcal{G}_{\sigma,p}$, and all $\phi \in S^1$. The adiabatic corrections $\tilde{F}$ and $\tilde{\Omega}$ are small in Gevrey spaces,*

$$|\tilde{F}(\tilde{u}; \varepsilon)|_{\mathcal{G}_{\sigma,p}} \leq C_1 \epsilon^{\frac{p}{p+1}}, \qquad |\tilde{\Omega}(\tilde{u}; \varepsilon)| \leq C_1 \epsilon,$$

*for $\tilde{u} \in B_R \subset \mathcal{G}_{\sigma,p}$.*

The proof of the theorem will be given in Section 3.1.

Note that until now we did not make any assumptions on well-posedness of the evolution equation. Indeed, a similar approach has previously been used to average ill-posed, elliptic problems, when rewritten as an evolution equation; see [Mat00]. Our theorem partially covers this case, also.

On the other hand, comparing solutions to (2.6) with solutions of the adiabatic approximation, that is, (2.6) with $\alpha, \beta$ set to zero, we need further information on the (unbounded) linear part $A$ in equation (2.1).

For $A$ sectorial, dissipative, the estimates from Theorem 1 imply that the fast phase can be adiabatically eliminated, causing an exponentially small error on finite time intervals; see [Mat01][Theorem 1, Theorem2] for a similar statement. Finite order estimates over infinite time-intervals have been pursued in [FiVi00], [Vis00] for dissipative evolution equations.

Corollary 1, below, shows that adiabatic elimination causes exponentially small errors for not necessarily dissipative, strongly continuous prinicipal parts $A$ of the evolution equation.

**Hypothesis 2.6** (Well-posedness). Assume $A$ is the infinitesimal generator of a strongly continuous semigroup on both $X$ and $\mathcal{G}_{\sigma,p}$.

By smoothness of the nonlinearities, the hypothesis implies that equation (2.1) possesses unique mild solutions in both $X$ and $\mathcal{G}_{\sigma,p}$; see Pazy [Paz83], for example.

We compare solutions of (2.6) with solutions of a truncated equation, with non-adiabatic terms neglected,

$$(2.8) \qquad \begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}u &= Au + F(u) + \tilde{F}(u; \varepsilon), \\ \frac{\mathrm{d}}{\mathrm{d}t}\phi &= \frac{1}{\varepsilon}(\Omega(u) + \tilde{\Omega}(u; \varepsilon)). \end{aligned}$$

**Corollary 1** (Gronwall estimates). *Under the assumptions of Theorem 1, assume, in addition, that Hypothesis 2.6 is satisfied. Fix $R > 0$, the maximal amplitude of the solution. Then for any $T > 0$ there are constants $\varepsilon_0(T) > 0$, and $C'(T), c'(T) > 0$ such that the following holds.*

*Let $u(t)$ be a solution to the truncated equation (2.8) with norm bounded by $R$ in the Gevrey $\mathcal{G}_{\sigma,p}$, for a time interval $0 \leq t \leq T < \infty$.*

*Then there exists a unique solution $(\tilde{u}(t), \phi(t))$ on $0 \leq t \leq T$ to (2.6) with initial value $(u(0), \phi_0)$ for any $\phi_0 \in S^1$. Moreover, the solutions are exponentially close in $\varepsilon < \varepsilon_0$,*

$$|\tilde{u}(t) - u(t)|_X \leq C' \exp\left(-c'\varepsilon^{-\frac{p}{1+p}}\right),$$

*for all $0 \leq t \leq T$.*

*Proof of Corollary 1.* The difference $w(t) = \tilde{u}(t) - u(t)$ satisfies the equation

$$\frac{\mathrm{d}}{\mathrm{d}t}w = A(t)w + r(t),$$

where

$$A(t) = A + \int_0^1 (\partial_u F + \partial_u \tilde{F})(\tau u + (1-\tau)\tilde{u})\mathrm{d}\tau$$

and

$$r(t) = \alpha(\tilde{u}(t)) - \alpha(u(t)),$$

where for the sake of notation, we omitted the arguments $\phi$ and $\varepsilon$. Since $F$, $\tilde{F}$, and $\alpha$ possess bounded derivatives on bounded sets of $\mathcal{G}_{\sigma,p}$, the result is an immediate consequence of the Gronwall Lemma. Note that the estimate on $\alpha$, equation (2.7) is in the $X$-topology, only, such that the closeness result only holds in this topology, as well. □

### 2.2. Averaging Hamiltonian evolution equations.
A particularly interesting setting, where rapid oscillations naturally occur, is Hamiltonian systems. There are many examples of Hamiltonian partial differential equations, and, as far as existence and regularity of solution is concerned, there are fundamental differences between the examples. We restrict here to a continuous symplectic form, which basically reduces to the standard symplectic form on finite-dimensional symplectic subspaces.

Let $Z$ be a possibly infinite-dimensional Hilbert space. Consider the phase space $(Z \times Z) \times (S^1 \times \mathbb{R})$, which we identify with the cotangent bundle of $Z \times S^1$. The standard symplectic form on this cotangent bundle is provided by

$$\begin{aligned}
(2.9) \quad & \omega\left((u_1, v_1, \psi_1, I_1), (u_2, v_2, \psi_2, I_2)\right) \\
& = (u_1, v_2)_Z - (u_2, v_1)_Z + \mathrm{d}\psi \wedge \mathrm{d}I\left((\psi_1, I_1), (\psi_2, I_2)\right).
\end{aligned}$$

Here, $(.,.)_Z$ denotes the inner product in $Z$ and $\mathrm{d}\psi \wedge \mathrm{d}I$ is the standard symplectic form on the cylinder $S^1 \times \mathbb{R}$ with coordinates $(\psi, I)$; see, e.g., [MaRa94].

Partial derivatives in the Hamiltonian equation are represented by a quadratic, closed form on the phase space.

Let $L : D(L) \subset Z \to Z$ be a selfadjoint, possibly unbounded linear operator, with domain dense in $Z$.

**Hypothesis 2.7** (Hamiltonian Galerkin approximation). We assume that there exists a sequence of orthogonal (Galerkin) projections $(Q_m)_{m \in \mathbb{N}}$ which satisfy the following requirements:

(i) the sequence of projections converges strongly to the identity on $Z$,

$$\lim_{m \to \infty} Q_m u = u \text{ in } Z \text{ for all } u \in Z;$$

(ii) the projections $Q_m$ commute with $L$ on its domain of definition

$$Q_m L u = L Q_m u \text{ for all } u \in D(L);$$

(iii) the operator $L$ is bounded on $\mathrm{Rg}\, Q_m$,

$$|LQ_m u|_Z \leq m|Q_m u|_Z \text{ for all } u \in Z;$$

(iv) on $\mathrm{Ker}\, Q_m$, we have the resolvent estimate

$$|(L|_{\mathrm{Ker}\, Q_m})^{-1}|_{\mathcal{L}(Z,Z)} \leq C_0/m,$$

with $C_0$ independent of $m$.

On the phase space, consider a Hamiltonian of the form

$$(2.10) \quad H(u, v, I, \psi) = \frac{1}{2} \left( (Lu, u)_Z + (Lv, v)_Z \right) + \frac{\Omega(I)}{\varepsilon} + H_0(u, v, I) + H_1(u, v, I, \psi),$$

where $I \in \mathbb{R}$, $\psi \in S^1$ and $u, v \in D(L)$. The nonlinear functions $\Omega : \mathbb{R} \to \mathbb{R}$, $H_0 : \mathbb{Z} \times \mathbb{Z} \times \mathbb{R} \to \mathbb{R}$, and $H_1 : \mathbb{Z} \times \mathbb{Z} \times \mathbb{R} \times S^1 \to \mathbb{R}$ will satisfy certain smoothness assumptions, which we will specify in Hypothesis 2.10, below.

Similarly to Hypothesis 2.2 and Hypothesis 2.3, we impose a condition on the frequency $\Omega$.

**Hypothesis 2.8** (High frequency oscillation). We assume that there is $c > 0$ such that $\Omega'(I) > c$ for all $I \in \mathbb{R}$.

Again, we assume that the non-adiabatic terms have zero average.

**Hypothesis 2.9** (Zero mean). We assume that $\int_{S^1} H_1(u, v, I, \psi) d\psi = 0$ for all $I \in \mathbb{R}$ and $u, v \in Z$.

The assumption corresponding to Hypothesis 2.4 was already encoded in Hypothesis 2.7,(iv).

From the Hamiltonian $H$ and the symplectic structure $\omega$, we derive the following, general system of abstract Hamiltonian evolution equations:

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \partial_v H \\ -\partial_u H \end{pmatrix} = \begin{pmatrix} Lv \\ -Lu \end{pmatrix} + \begin{pmatrix} \partial_v(H_0 + H_1) \\ -\partial_u(H_0 + H_1) \end{pmatrix},$$

$$(2.11) \quad \frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} \psi \\ I \end{pmatrix} = \begin{pmatrix} \partial_I H \\ -\partial_\psi H \end{pmatrix} = \begin{pmatrix} \frac{\Omega'(I)}{\varepsilon} \\ 0 \end{pmatrix} + \begin{pmatrix} \partial_I(H_0 + H_1) \\ -\partial_\psi(H_1) \end{pmatrix}.$$

The unbounded operator $A : D(A) = D(L) \times D(L) \to Z \times Z =: X$ appearing on the right side, defined through

$$A \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} Lv \\ -Lu \end{pmatrix},$$

is a closed, densely defined, anti-symmetric operator. It satisfies Hypothesis 2.1, when choosing $P_m := \mathrm{diag}\,(Q_m, Q_m)$.

As in the previous section, we define the Gevrey classes $\mathcal{G}_{\sigma,p}$ by means of the exponential $\Gamma_{\sigma,p} = \exp(\sigma|A|^p)$. Note that $\mathcal{G}_{\sigma,p} \subset Z \times Z$ is dense.

The smoothness assumptions on the non-quadratic part of the Hamiltonian are:

**Hypothesis 2.10** (Analyticity of nonlinearities). There is a Gevrey class $Y = \mathcal{G}_{\sigma,p}$ with $\sigma, p > 0$ and a constant $\delta > 0$ for the size of the complex extension such that the following properties of the non-quadratic part of the Hamiltonian hold.

The nonlinearities are analytic and bounded on bounded subsets when considered on Gevrey spaces, extended in the complex direction,

$$H_0 : (Y + \delta) \times (\mathbb{R} + \delta) \to \mathbb{C} \quad \Omega : (\mathbb{R} + \delta) \to \mathbb{C},$$
$$H_1 : (Y + \delta) \times (\mathbb{R} + \delta) \times (S^1 + \delta) \times (\mathbb{R} + \delta) \to \mathbb{C}.$$

In addition, the above statements are assumed to hold when the space of Gevrey regularity $Y = \mathcal{G}_{\sigma,p}$ is replaced by $Y := X = Z \times Z$.

As in Section 2.1, we are now looking for a change of coordinate which would transform the equation with fast and slow motion decoupled in the new coordinates. Only the $\psi$-variable is fast and, therefore, the main obstacle is the dependence of $H_1$

on $\psi$. To find such coordinates, we might simply invoke Theorem 1. However, the change of coordinate might not respect the symplectic structure, and the system might not be Hamiltonian with respect to the standard symplectic structure $\omega$.

**Definition 2.11** (Canonical change of coordinates). A diffeomorphism $\mathcal{W} : \mathcal{U} \subset Z \times Z \times S^1 \times \mathbb{R} \to \mathcal{V} \subset Z \times Z \times S^1 \times \mathbb{R}$ is said to be canonical if it preserves the symplectic form

$$\omega(\mathcal{W}'\cdot, \mathcal{W}'\cdot) = \omega(\cdot, \cdot).$$

We are now ready to state our second main result.

**Theorem 2.** *Assume Hypothesis 2.7 on the existence of Hamiltonian Galerkin approximations, Hypothesis 2.8 for a minimal large oscillation frequency, and Hypothesis 2.10 on analyticity of the non-quadratic part of the Hamiltonian in some Gevrey class $\mathcal{G}_{\sigma,p} + \delta$. Then for any $R > 0$, there exists a canonical transformation $\mathcal{W} = \mathrm{id} + \varepsilon W$, defined on $B_R \subset Z \times Z \times S^1 \times \mathbb{R}$, which eliminates adiabatically the fast phase, up to an exponentially small non-adiabatic effect.*

*More precisely, the transformation is analytic on balls $B_R$ in the complex extensions $(\mathcal{G}_{\sigma,p} + \delta') \times (S^1 + \delta') \times (\mathbb{R} + \delta')$ and $(X + \delta') \times (S^1 + \delta') \times (\mathbb{R} + \delta')$, for some $\delta' > 0$.*

*In the new variables $(\tilde{u}, \tilde{v}, J, \phi)$, the transformed Hamiltonian reads*

$$(2.12) \qquad \begin{aligned} \tilde{H}(\tilde{u}, \tilde{v}, J, \phi) = {} & \frac{1}{2} \left( (L\tilde{u}, \tilde{u}) + (L\tilde{v}, \tilde{v}) \right) + \frac{\Omega(J)}{\varepsilon} + H_0(\tilde{u}, \tilde{v}, J) \\ & + \tilde{H}_1(\tilde{u}, \tilde{v}, J; \varepsilon) + H_r(\tilde{u}, \tilde{v}, J, \psi; \varepsilon). \end{aligned}$$

*The adiabatic correction $\tilde{H}_1$ of the Hamiltonian is $O(\varepsilon)$ and $H_r$ is bounded on bounded subsets of $Z \times Z \times \mathbb{R}$ and $Z \times Z \times \mathbb{R} \times S^1$ respectively.*

*The non-adiabatic interaction term $H_r$ is exponentially small in $\varepsilon$ when considered on the Gevrey class $\mathcal{G}_{\sigma,p}$. More precisely, there exist constants $c_1, C_1 > 0$ such that*

$$(2.13) \qquad |\tilde{H}_1(\tilde{u}, \tilde{v}, \phi, J; \varepsilon)| < C_1 \exp\left( -c_1 \varepsilon^{-\frac{p}{1+p}} \right),$$

*for all $(\tilde{u}, \tilde{v}) \in B_r \subset \mathcal{G}_{\sigma,p}$, $|J| \le R$, $\phi \in S^1$, and $\varepsilon > 0$ small.*

The proof of the theorem will be given in Section 3.2.

A result similar to Corollary 1 holds in the context of Hamiltonian systems.

## 2.3. Comments, related results, and open problems.

Many generalizations of our two main theorems are possible. First, continuous dependence on $\varepsilon$ is sufficient, as long as analyticity bounds are uniform in $\varepsilon > 0$. On the other hand, if the equation is smooth in $\varepsilon$, then the change of coordinates can be chosen to depend smoothly on $\varepsilon$. Dependence on $\psi$ can be relaxed in the particular case when $\dot{\psi} = 1/\varepsilon$, that is, when $\psi$ merely describes fast time.

Boundedness of the nonlinearities on the entire function space is rather unrealistic. However, the theorems still hold on bounded subsets of $X$ and $\mathcal{G}_{\sigma,p}$, when the assumption on boundedness in Hypotheses 2.5 and 2.10 on analyticity is replaced by boundedness on bounded subsets of $X$ and $\mathcal{G}_{\sigma,p}$, respectively.

In the Hamiltonian case, it is sometimes useful to consider a slightly more general setup. Whereas the symplectic structure and the equation might be defined on $X$, typically an energy space like $L^2$, the nonlinearities often are analytic only on spaces of higher regularity. Theorem 2 generalizes easily to the case when

the nonlinearities satisfy Hypothesis 2.10 on a fractional power interpolation space $X^s = D(|A|^s)$ equipped with the graph norm, rather than considered on the entire function space $X$. Then it is convenient also to consider modified Gevrey spaces $\mathcal{G}^s_{\sigma,p} = D(|A|^s \exp(\sigma|A|^p))$, as well.

The original equations might possess structure other than Hamiltonian. For example, the system might be invariant under the action of a finite-dimensional Lie group $\Gamma$; see, for example, [SaScWu99] for a general setup of equivariance and symmetry in the context of semilinear evolution equations. The averaging transformation would then preserve this symmetry of the system. A particular case of symmetry occurs in the Hamiltonian setup. Symmetries correspond, by Noether's theorem, to conserved quantities. Symmetries are preserved throughout the averaging procedure, which is equivalent to respecting preserved quantities. Note, however, that the adiabatically truncated equation (2.8) might possess additional (or less!) symmetries than the original or the transformed system (2.1), (2.6). Theorem 2 does not immediately imply a Nekhoroshev type stability theorem [Nek79] in the case when the adiabatically truncated system is completely integrable; see, however, [Bam99], [Poe99] for results in this direction.

An important conserved quantity in Hamiltonian PDEs is momentum, corresponding to spatial translational invariance of the original PDE. The averaging procedure respects this symmetry. However, as a drawback, we are not able to guarantee that the transformed equation is again *local* in the sense that nonlinearities only involve point evaluations. In general, averaging PDEs introduces non-local spatial averages already at low orders in $\varepsilon$. A similar phenomenon is well-known in reaction-diffusion equations, where rapidly diffusing species can be (formally) adiabatically eliminated at the expense of non-local coupling terms in the slow species. We are not aware of results showing that non-local equations evoke phenomena which are not present in local PDEs.

The small non-adiabatic interaction terms have in general nonzero average over $S^1$. The theorems remain valid, if we impose zero mean, but then the estimates on the adiabatic corrections $\tilde{F}$ and $\tilde{\Omega}$ only provide boundedness.

We believe that the regularity assumptions on the initial data are crucial. In the case of parabolic equations, this assumption can be dropped, since, after a short initial transient, the parabolic smoothing regularizes the initial data such that the solution belongs to Gevrey spaces with $p = 1/2$; see [Mat01]. Even (ill-posed) elliptic equations on infinite cylinders can be formulated as in (2.1) and exponential averaging applies with $p = 1$; see [Mat00].

The simplest example is the case of bounded $A$. Refining the proof below slightly, it is then not difficult to show that we can obtain $p = \infty$, that is, exponential averaging with exponent $-c\varepsilon^{-1}$ as in the case of ODEs, see [Nei84], is possible.

## 3. Proofs of Theorems 1 and 2

We prove our main results on exponential smallness of non-adiabatic interaction in infinite-dimensional evolution systems. In Section 3.1, we prove Theorem 1 for general evolution equations, and in Section 3.2, we sketch the necessary adaptions in the Hamiltonian setup, in order to prove Theorem 2.

3.1. **Proof of Theorem 1.** We outline the proof. The proof consists of two main parts. We split the system by means of a Galerkin projection $P_N$ from Hypothesis 2.1 into a system where the unbounded operator is bounded, and a complement.

In the bounded part $u \in X_N := P_N X$, we perform successive averaging steps inspired by Neishtadt's proof in the case of ordinary differential equations. The main difficulty turns out to preserve the semilinear structure of the equation despite the norm of the vector field growing with increasing accuracy $N$ of the Galerkin approximations $P_N$. The key idea is to choose $N$, which is also the norm of $A$ on the Galerkin approximation space $P_N X$, proportional to a power of the period $\varepsilon$ of the forcing, $N \sim \varepsilon^{\rho'}$. The exponent $\rho'$ is yet to be determined. In other words, the higher the frequency, the more accurate we can choose our approximation to the original PDE.

Inside this (possibly infinite-dimensional) subspace, we construct a transformation, such that the non-adiabatic remainder is exponentially small in the frequency $\varepsilon$. This is achieved by a sequence of successive transfomations, each of which decreases the contribution of the non-adiabatic term by a constant factor. The number of consecutive transformations is chosen as a function of $\varepsilon$, as well. The exponent of the frequency $\varepsilon$ in the exponentially small non-adiabatic contribution within the Galerkin approximation still depends on the exponent in the coupling between the accuracy of the Galerkin approximation $N$ and the frequency $\varepsilon$. An optimal choice of this exponent is provided by comparison with possible non-adiabatic terms created through coupling to spatially small-scale structures in the complement of $\mathrm{Rg}\,(P_N)$. These terms appear in the second part of the proof, when extending the transformation to the full system $u \in X$. Instead of attempting to average these contributions from the dynamics in the small spatial scales, we estimate the overall contribution exploiting the high spatial regularity. Indeed, Gevrey regularity guarantees that the contributions from small spatial scales are indeed exponentially small in $N$. Balancing the exponential estimates for the averaging procedure in $\mathrm{Rg}\,(P_N)$, Step 1, and the overall contribution of small spatial scales, Step 2, we find an optimal exponent $\rho$ for the coupling between temporal frequency $\varepsilon$ and spatial resolution $N$. This yields the estimate on the non-adiabatic terms in Theorem 1.

*Step* 1. *Averaging the Galerkin approximation.* We approximate the evolution equation (2.1), by means of the Galerkin projections $P_N$ from Hypothesis 2.1, with an equation on $X_N := \mathrm{Rg}\,(P_N)$ coupled to the fast phase $\psi$:

$$
\begin{aligned}
\dot{u}_N &= Au_N + P_N F(u_N) + P_N G(u_N, \psi; \varepsilon), \\
\dot{\psi} &= \frac{\Omega(u_N)}{\varepsilon} + g(u_N, \psi; \varepsilon).
\end{aligned}
$$

(3.1)

We couple the scale of the spatial resolution $N^{-1}$ and the temporal frequency $\varepsilon$, choosing $N \in \mathbb{N}$ maximal such that

$$
(3.2) \qquad\qquad \varepsilon \le N^{-(1+\rho)}.
$$

The exponent $\rho$ will be chosen in Step 2, when comparing the estimate from the averaging procedure below to the Gevrey estimates on small spatial scales.

We perform a sequence of successive near-identity transformations

$$
\begin{pmatrix} u_N \\ \psi \end{pmatrix} = \begin{pmatrix} \tilde{u}_N \\ \phi \end{pmatrix} + \varepsilon W_k \left( \tilde{u}_N, \phi; \varepsilon \right).
$$

The total transformation after $k$ steps is given by the composition of the transformations $(\mathrm{id} + W_k) \circ (\mathrm{id} + W_{k-1}) \circ \ldots \circ (\mathrm{id} + W_1)$. After $k$ transformations, the

equation in the new variable $\tilde{u}$ can be written in the general form

$$
\begin{aligned}
\dot{u}_N &= Au_N + P_N F(u_N) + \tilde{F}_k(u_N; \varepsilon) + \alpha_k(u_N, \psi; \varepsilon), \\
\dot{\psi} &= \frac{\Omega_k(u_N; \varepsilon)}{\varepsilon} + \beta_k(u_N, \psi; \varepsilon).
\end{aligned}
$$

(3.3)

We may assume that the average of both $\alpha_k$ and $\beta_k$ over the phase $\psi \in S^1$ vanishes, possibly changing $\tilde{F}_k$ and $\Omega_k$:

$$
\langle \alpha_k(u_N, \cdot; \varepsilon) \rangle := \int_{S^1} \alpha_k(u_N, \psi; \varepsilon) \mathrm{d}\psi = 0,
$$

$$
\langle \beta_k(u_N, \cdot; \varepsilon) \rangle := \int_{S^1} \beta_k(u_N, \psi; \varepsilon) \mathrm{d}\psi = 0.
$$

By Hypothesis 2.3, the terms $G$ and $g$ from the original equation fulfill this assumption. Comparing equation (3.1) with (3.3), we find the following expressions for the terms in (3.3) before the first transformation

$$
\tilde{F}_0(u; \varepsilon) = 0, \quad \alpha_0(u; \varepsilon) = P_N G(u, \psi; \varepsilon),
$$

(3.4)

and

$$
\Omega_0(u; \varepsilon) = \Omega(u), \quad \beta_0(u, \psi; \varepsilon) = g(u, \psi; \varepsilon).
$$

(3.5)

For the $(k+1)$-th transformation, $k = 0, 1, 2, \ldots$, we choose the explicit change of coordinate

$$
W_{k+1}(\tilde{u}_N, \phi; \varepsilon) = \begin{pmatrix} W^1_{k+1}(\tilde{u}_N, \phi; \varepsilon) \\ W^2_{k+1}(\tilde{u}_N, \phi; \varepsilon) \end{pmatrix},
$$

(3.6)

with

$$
W^1_{k+1}(\tilde{u}_N, \phi; \varepsilon) = \frac{1}{\Omega_k(u_N; \varepsilon)} \int_0^\phi \alpha_k(\tilde{u}_N, \tau; \varepsilon) \mathrm{d}\tau
$$

and

$$
\begin{aligned}
W^2_{k+1}(\tilde{u}_N, \phi; \varepsilon) = \frac{1}{\Omega_k(\tilde{u}_N; \varepsilon)} \int_0^\phi &\beta_k(\tilde{u}_N, \tau; \varepsilon) \\
&+ \partial_{\tilde{u}_N} \Omega_k(\tilde{u}_N; \varepsilon)(W^1_{k+1}(\tilde{u}_N, \tau; \varepsilon) - \langle W^1_{k+1}(\tilde{u}_N, \cdot; \varepsilon) \rangle) \mathrm{d}\tau,
\end{aligned}
$$

where $\langle . \rangle$ again denotes the $S^1$-average. Since the averages of $\alpha_k$ and $\beta_k$ vanish, the transformation is well defined for $\phi \in S^1$. This transformation can formally be viewed as the classical averaging transformation, which eliminates dependence on the fast phase $\psi$ in $\alpha_k$ and $\beta_k$ to leading order. Substituting the transformation

into (3.3) and writing $u$, $P$ instead of $\tilde{u}_N$, $P_N$ for the sake of notation yields

$$
\frac{\mathrm{d}u}{\mathrm{d}t} = \left(\mathrm{id} + \varepsilon\partial_u W_{k+1}^1(u, \phi; \varepsilon)\right)^{-1}
$$

$$
\Big\{ A(u + \varepsilon W_{k+1}^1(u, \phi; \varepsilon)) + PF(u + \varepsilon W_{k+1}^1(u, \phi; \varepsilon))
$$

$$
+\tilde{F}_k(u + \varepsilon W_{k+1}^1(u, \phi; \varepsilon); \varepsilon) + \alpha_k(u + \varepsilon W_{k+1}^1(u, \phi; \varepsilon), \psi; \varepsilon)
$$

$$
-\partial_\phi W_{k+1}^1(u, \phi; \varepsilon)[\Omega_k(u + \varepsilon W_{k+1}^1(u, \phi; \varepsilon); \varepsilon)]
$$

$$
-\partial_\phi W_{k+1}^1(u, \phi; \varepsilon)[\varepsilon\beta_k(u + \varepsilon W_{k+1}^1(u, \phi; \varepsilon), \phi; \varepsilon)] \Big\}
$$

$$
(3.7) \qquad =: \quad Au + PF(u) + \tilde{F}_k(u; \varepsilon) + a(u, \phi; \varepsilon),
$$

$$
\frac{\mathrm{d}\phi}{\mathrm{d}t} = \left(1 + \varepsilon\partial_\phi W_{k+1}^2(u, \phi; \varepsilon)\right)^{-1}
$$

$$
\Big\{ \frac{\Omega_k(u + \varepsilon W_{k+1}^1(u, \phi; \varepsilon); \varepsilon)}{\varepsilon} + \beta_k(u + \varepsilon W_{k+1}^1(u, \phi; \varepsilon), \phi; \varepsilon)
$$

$$
-\varepsilon\partial_u W_2(u, \phi; \varepsilon)\frac{\mathrm{d}u}{\mathrm{d}t} \Big\}
$$

$$
(3.8) \qquad =: \quad \frac{\Omega_k(u; \varepsilon)}{\varepsilon} + \partial_u\Omega_k(u; \varepsilon)\langle W_{k+1}^1(u, \cdot; \varepsilon)\rangle + b(u, \phi; \varepsilon).
$$

Here, we singled out the new correction terms $a, b$ appearing in the equation. In order to write the system in the inductive form (3.2), we define

$$
\begin{aligned}
\tilde{F}_{k+1}(u; \varepsilon) &:= & \tilde{F}_k(u; \varepsilon) + \langle a(u, \cdot; \varepsilon)\rangle, \\
\Omega_{k+1}(u; \varepsilon) &:= & \Omega_k(u; \varepsilon) + \varepsilon\partial_u\Omega_k(u; \varepsilon)\langle W_{k+1}^1\rangle + \varepsilon\langle b(u, \cdot; \varepsilon)\rangle, \\
\alpha_{k+1}(u, \psi; \varepsilon) &= & a(u, \psi; \varepsilon) - \langle a(u, \cdot; \varepsilon)\rangle, \\
(3.9) \qquad \beta_{k+1}(u, \psi; \varepsilon) &= & b(u, \psi; \varepsilon) - \langle b(u, \cdot; \varepsilon)\rangle.
\end{aligned}
$$

Our objective now is to show that the norms of the non-adiabatic terms $\alpha_k$ and $\beta_k$ are decreasing in $k$. In fact, we will show that in suitably chosen norms, the size is reduced by at least a factor 2 in each step. The main ingredient is the analytic extension of functions into the complexified space. Actually, all functions appearing in the above equations are analytic on the Gevrey space $\mathcal{G}_{\sigma,p}$ by Hypothesis 2.5. Note that the linear projection $P = P_N$ is analytic, as well.

We need analyticity to compensate for the loss of derivatives in the averaging steps. Basically we rely on a variant of the Cauchy estimate. Recall the notation $Y + \delta \subset Y_{\mathbb{C}}$ for the complex $\delta$-extension of the real subspace $Y \subset Y_{\mathbb{C}}$, as defined in (2.5).

**Lemma 3.1** (Cauchy estimate). *Let $f : (Y + \delta) \to Y_{\mathbb{C}}$ be an analytic function with $\|f\|_{Y+\delta} := \sup_{u \in (Y+\delta)} \|f(u)\| < \infty$. Then for all $0 < \eta < \delta$,*

$$
\|\partial_u f\|_{Y+(\delta-\eta)} \leq \frac{1}{\eta}\|f\|_{Y+\delta}.
$$

The proof is a simple application of the Cauchy formula; see, e.g., [Mat01, Lemma 7].

*The induction assumption and exponential estimates.* We are now going to show inductively that for all $1 \leq k \leq K(\varepsilon)$ and some constant $M$, we can estimate the

nonlinearities in the transformed equation (3.3) by

$$(3.10) \qquad \|\tilde{F}_k\|_{\mathcal{Y}+\delta_k} \;\leq\; 2M\varepsilon^{\frac{\rho}{1+\rho}},$$

$$(3.11) \qquad \|\alpha_k\|_{\mathcal{Y}+\delta_k} \;\leq\; 2^{-k}M\varepsilon^{\frac{\rho}{1+\rho}},$$

$$(3.12) \qquad |\beta_k|_{\mathcal{Y}+\delta_k} \;\leq\; 2^{-k}M\varepsilon^{\frac{\rho}{1+\rho}},$$

$$(3.13) \qquad |\Omega - \Omega_k|_{\mathcal{Y}+\delta_k} \;\leq\; M\varepsilon \text{ and } \Omega_k > \frac{c}{2}.$$

Here, $\mathcal{Y} = P_N \mathcal{G}_{\sigma,p}$,

$$(3.14) \qquad \delta_k = \delta - k\eta(\varepsilon), \qquad \eta(\varepsilon) = \eta_0 \varepsilon^{\frac{\rho}{1+\rho}},$$

$c$ is the positive constant from Hypothesis 2.2, and the norms on the left side are defined by

$$\|\tilde{F}_k\|_{\mathcal{Y}+\delta_k} := \sup\{\|\tilde{F}_k(u;\varepsilon)\|_{\mathcal{G}_{\sigma,p}};\ u \in \mathcal{Y} + \delta_k\},$$

and analogous expressions for $\alpha_k$, $\beta_k$, and $\Omega - \Omega_k$. The maximal number of steps $K(\varepsilon)$ is given by the constraint that the domain $\mathcal{Y} + \delta_k$ should be nonempty, that is, $\delta_{K(\varepsilon)} = \delta - K(\varepsilon)\eta_0 \varepsilon^{\frac{\rho}{1+\rho}} > 0$, which gives

$$(3.15) \qquad K(\varepsilon) = \frac{\delta}{\eta_0}\varepsilon^{-\frac{\rho}{1+\rho}}.$$

Assume for the moment, the above inequalities (3.10)–(3.13) hold. Choosing $K(\varepsilon)$ maximal such that the $\delta_k$ do not vanish, (3.15), we find exponential estimates for the non-adiabatic interaction terms

$$(3.16) \qquad \|\alpha_{K(\varepsilon)}\|_{\mathcal{Y}} + |\beta_{K(\varepsilon)}|_{\mathcal{Y}} \leq 2M\varepsilon^{-\frac{\rho}{1+\rho}}2^{-K(\varepsilon)} \leq C_0 \mathrm{e}^{-c_0\varepsilon^{-\frac{\rho}{1+\rho}}}.$$

Therefore, in order to show exponential averaging within the Galerkin approximation, it is sufficient to show inductively that the inequalities (3.10)–(3.13) are satisfied.

For $k = 1$, the assertions follow from a direct estimate, which we postpone until after the induction step.

*The induction step.* Assume the inductive assumptions (3.10)–(3.13) are fulfilled for the first $k$ steps.

*Notation.* To simplify notation we write $Y_k = \mathcal{Y} + \delta_k$, $M_k = 2^{-k}M\varepsilon^{\frac{\rho}{\rho+1}}$, and suppress arguments of $W_k^j$ and dependence on the parameter $\varepsilon$ whenever possible. We denote by $C$ a universal constant, which does not depend on $k$ and $\varepsilon$. Note that we write norms on the function spaces $X$, $\mathcal{G}_{\sigma,p}$, etc., although the functions are only defined on balls of radius $R$, arbitrarily large.

We start providing estimates on the $(k+1)$-th coordinate change. By definition, equation (3.6), we obtain

$$(3.17) \qquad \|\varepsilon W_{k+1}^1\|_{Y_k} \leq M_k\varepsilon, \qquad |\varepsilon W_{k+1}^2|_{Y_k} \leq 2MM_k\varepsilon.$$

With the Cauchy estimate, Lemma 3.1, we conclude that

$$(3.18) \qquad \|\varepsilon\partial_u W_{k+1}^1\|_{Y_{k+1}} \leq \frac{M_k\varepsilon}{\eta(\varepsilon)}.$$

From the definition of $W_{k+1}^2$, and directly from (3.6), we find

$$(3.19) \qquad \|\varepsilon\partial_\phi W_{k+1}^2\|_{Y_{k+1}} \leq \varepsilon\frac{2}{c}\left(M_k + \frac{M}{\eta(\varepsilon)}M_k\right) \leq \frac{3\varepsilon MM_k}{c\eta(\varepsilon)}.$$

The estimates on $\alpha_{k+1}$ and $\tilde{F}_{k+1}$, (3.10) and (3.11), are derived from an estimate on the higher-order term $a$ in (3.7). Inspecting the definition of $a$ and rearranging terms we find

$$
\begin{aligned}
a &= \left(I + \varepsilon \partial_u W_{k+1}^1\right)^{-1} \Big\{ PF(u + \varepsilon W_{k+1}^1) + \tilde{F}_k(u + \varepsilon W_{k+1}^1) + \alpha_k(u + \varepsilon W_{k+1}^1) \\
&\qquad - \frac{\alpha_k(u)}{\Omega_k(u)} \left(\Omega_k(u + \varepsilon W_{k+1}^1) + \varepsilon \beta_k(u + \varepsilon W_{k+1}^1)\right) + A(u + \varepsilon W_{k+1}^1) \Big\} \\
&\quad - \Big[ Au + PF(u) + \tilde{F}_k(u) \Big] \\
&= \left(I + \varepsilon \partial_u W_{k+1}^1\right)^{-1} \Big\{ \varepsilon A W_{k+1}^1 + PF(u + \varepsilon W_{k+1}^1) - PF(u) + \tilde{F}_k(u + \varepsilon W_{k+1}^1) \\
&\qquad - \tilde{F}_k(u) + \alpha_k(u + \varepsilon W_{k+1}^1) - \frac{\alpha_k(u)}{\Omega_k(u)} \left(\Omega_k(u + \varepsilon W_{k+1}^1) + \varepsilon \beta_k(u + \varepsilon W_{k+1}^1)\right) \\
&\qquad\qquad\qquad\qquad\qquad - \varepsilon \partial_u W_{k+1}^1 \Big[ Au + PF(u) + \tilde{F}_k(u) \Big] \Big\}.
\end{aligned}
$$

Estimating the norms in $Y_{k+1}$, using $\|Au\| \le N\|u\|$ yields

$$
\begin{aligned}
\|a\| &\le \left(1 - \|\varepsilon \partial_u W_{k+1}^1\|\right)^{-1} \\
&\quad \cdot \Big\{ \|\varepsilon W_{k+1}^1\| \Big[ N + \|\partial_u PF\| + \|\partial_u \tilde{F}_k\| + \|\partial_u \alpha_k\| + C|\partial_u \Omega_k(u)| \Big] \\
&\qquad + \left| \frac{\varepsilon \alpha_k(u) \beta_k(u + \varepsilon W_{k+1}^1)}{\Omega_k(u)} \right| + \|\varepsilon \partial_u W_{k+1}^1\| \Big[ N\|u\| + \|PF\| + \|\tilde{F}_k\| \Big] \Big\}.
\end{aligned}
$$

(3.20)

Together with the Cauchy estimate, Lemma 3.1, the a priori bound $\|u\| \le R$, and the induction assumption, we can estimate the norms as follows:

$$
\begin{aligned}
\|a\|_{Y_{k+1}} &\le 2\Big\{ \varepsilon M_k \Big[ N + \frac{M}{\eta(\varepsilon)} + \frac{2M}{\eta(\varepsilon)} + \frac{CM}{\eta(\varepsilon)} + \frac{M_k}{\eta(\varepsilon)} \Big] \\
&\qquad + C\varepsilon M_k^2 + \frac{\varepsilon M_k}{\eta(\varepsilon)} [NR + M + 2M] \Big\}.
\end{aligned}
$$

Recall that $N \le \varepsilon^{-\frac{1}{1+\rho}}$, equation (3.2), and $\eta(\varepsilon) = \eta_0 \varepsilon^{\frac{\rho}{1+\rho}}$. With these appropriate scalings, we obtain

$$
\|a\|_{Y_{k+1}} \le M_k \left\{ \left[ 2\varepsilon^{\frac{\rho}{1+\rho}} + \frac{CM}{\eta_0} \varepsilon^{\frac{1}{1+\rho}} \right] + C\varepsilon M_k + \frac{R}{\eta_0} \right\} \le \frac{1}{4} M_k
$$

for $R/\eta_0 < 1/4$ and $\varepsilon$ small enough. Therefore,

$$
\|\alpha_{k+1}\|_{Y_{k+1}} < \frac{M_k}{2} = M_{k+1}
$$

and

$$
\|\tilde{F}_{k+1} - \tilde{F}_k\|_{Y^{k+1}} < \frac{M_k}{4}.
$$

This proves the first part of the induction step, the estimates (3.10) and (3.11) on $\alpha_k$ and $\tilde{F}_k$, with the given appropriate choice of $\eta(\varepsilon)$, (3.14).

We next address the estimates (3.12) and (3.13). We follow the same strategy and start computing the correction $b$:

$$
\begin{aligned}
b &= \left(1 + \varepsilon \partial_\phi W_{k+1}^2\right)^{-1} \left\{ \frac{\Omega_k(u + \varepsilon W_{k+1}^1)}{\varepsilon} + \beta_k(u + \varepsilon W_{k+1}^1) - \varepsilon \partial_u W_{k+1}^2 \frac{du}{dt} \right\} \\
&\quad - \frac{\Omega_k(u)}{\varepsilon} - \partial_u \Omega_k(u) \langle W_{k+1}^1 \rangle.
\end{aligned}
$$

We now take the modulus and estimate, again exploiting the Cauchy estimate, in $Y_{k+1}$. Expanding the first term as a series in $\varepsilon$, we find

$$|b| = \left| \left( 1 - \varepsilon \partial_\phi W_{k+1}^2 + O\left( \frac{M_k^2 \varepsilon^2}{\eta(\varepsilon)^2} \right) \right) \left\{ \frac{\Omega_k(u + \varepsilon W_{k+1}^1)}{\varepsilon} + \beta_k(u + \varepsilon W_{k+1}^1) \right. \right.$$
$$\left. \left. + O(\frac{\varepsilon N R}{\eta(\varepsilon)}) \right\} - \frac{\Omega_k(u)}{\varepsilon} - \partial_u \Omega_k(u) \langle W_{k+1}^1 \rangle \right|.$$

We next exploit the explicit form of the averaging transformation

$$\varepsilon \partial_\phi W_{k+1}^2 = \frac{1}{\Omega_k} \left( \beta_k + \partial_u \Omega(W_{k+1}^1 - \langle W_{k+1}^1 \rangle) \right),$$

to see, after a short computation, that again the leading-order terms drop out. The remainder terms are estimated as above.

After tedious but straightforward estimates we arrive at

$$|b| \leq \frac{1}{4} M_k,$$

for $\varepsilon_0$ small enough, and $\eta_0$ large. Consequently, we conclude that

$$\begin{aligned} |\beta_{k+1}| &< \frac{M_k}{2} = M_{k+1}, \\ |\Omega_{k+1} - \Omega_k| &< 2 M_k \varepsilon, \end{aligned}$$

and

$$|\Omega_{k+1}| > \frac{c}{2},$$

uniformly in $k$ and $\varepsilon$ small. This proves the remaining statements (3.12) and (3.13) of the induction procedure.

It remains to check the assertions for $k = 1$. Starting with (3.4), (3.5) we perform a first averaging step of the same form as in the induction step, only with a slightly different estimate. We use a fixed shrinking $\eta$ of the complex neighborhood $\mathcal{Y} + \delta$, instead of the $\varepsilon$-dependent one above. This allows us to estimate $a$ and $b$ as above and get

$$\begin{aligned} \|a\| &\leq M \varepsilon N = M \varepsilon^{\frac{\rho}{\rho+1}}, \\ \|b\| &\leq M \varepsilon N = M \varepsilon^{\frac{\rho}{\rho+1}}. \end{aligned}$$

This yields, possibly increasing $M$, the estimates (3.10)–(3.13) for $k = 1$.

Summarizing, we have shown that the Galerkin approximation with $N \leq N(\varepsilon)$, equation (3.2), can be averaged by a sequence of $K(\varepsilon)$ successive coordinate changes, with $K(\varepsilon)$ defined in (3.15), such that the resulting equation takes the form

$$\begin{aligned} \dot{u} &= Au + PF(u) + \tilde{F}_*(u; \varepsilon) + \alpha_*(u, \phi; \varepsilon), \\ (3.21) \qquad \dot{\phi} &= \frac{\Omega_*(u; \varepsilon)}{\varepsilon} + \beta_*(u, \phi; \varepsilon), \end{aligned}$$

with

$$(3.22) \qquad \sup_{|Pu|_{\mathcal{G}_{\sigma,p}}<R,\phi\in S^1} |\alpha_*(u,\phi;\varepsilon)|_{\mathcal{G}_{\sigma,p}} \quad < \quad C\exp\left(-c_2\varepsilon^{-\frac{\rho}{1+\rho}}\right),$$

$$(3.23) \qquad \sup_{|Pu|_{\mathcal{G}_{\sigma,p}}<R,\phi\in S^1} |\beta_*(u,\phi;\varepsilon)| \quad < \quad C\exp\left(-c_2\varepsilon^{-\frac{\rho}{1+\rho}}\right),$$

$$(3.24) \qquad \sup_{|Pu|_{\mathcal{G}_{\sigma,p}}<R} |\tilde{F}_*(u;\varepsilon)|_{\mathcal{G}_{\sigma,p}} \quad < \quad 2M\varepsilon^{\frac{\rho}{1+\rho}},$$

$$(3.25) \qquad \sup_{|Pu|_{\mathcal{G}_{\sigma,p}}<R} |\Omega_*(u;\varepsilon)-\Omega(u)| \quad < \quad 4M\varepsilon$$

uniformly in $0 < \varepsilon < \varepsilon_0$ with $N(\varepsilon)$ chosen by (3.2). Using exactly the same estimates just with the $\mathcal{G}_{\sigma,p}$-norm replaced by the $X$-norm, gives that $\tilde{F}_*$, $\alpha_*$, $\Omega_*/\varepsilon$, $\beta_*$ and $W$ are bounded on bounded sets in $X$. Note that the parameter $\rho > 0$ is still free. We will choose $\rho$ when taking into account the corrections to the Galerkin approximation in the next paragraph.

*Step* 2. *Averaging the full system.* The last step of the proof of Theorem 1 consists of extending the transformation to the full system and estimating the "small-scale" contributions from $\mathrm{Ker}\,(P_N)$.

We define the transformation on the entire space $X = \mathrm{Rg}\,(P_N) \oplus \mathrm{Ker}\,(P_N)$ through

$$u = \tilde{u} + \varepsilon W(P_N\tilde{u},\phi;\varepsilon,N),$$

where $W$ is the transformation, constructed in Step 1, with range contained in the not necessarily finite-dimensional image $\mathrm{Rg}\,(P_N)$. The order of approximation $N = N(\varepsilon)$ is chosen according to Step 1, equation (3.2), maximally, such that $N^{1+\rho}\varepsilon \leq 1$ holds. Substituting the transformation into the full system, equation (2.1), we find a transformed equation of the general form in (2.6),

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{u} \quad = \quad A\tilde{u} + F(\tilde{u}) + \tilde{F}(\tilde{u};\varepsilon) + \alpha(\tilde{u},\psi;\varepsilon),$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi \quad = \quad \frac{\tilde{\Omega}(\tilde{u};\varepsilon)}{\varepsilon} + \beta(\tilde{u},\phi;\varepsilon),$$

where

$$
\begin{aligned}
\tilde{F}(\tilde{u};\varepsilon) &= \tilde{F}_*(\tilde{u};\varepsilon), \\
\alpha &= \alpha_* + b_1 + b_2, \\
\tilde{\Omega} &= (\Omega - \Omega_*) + \varepsilon\langle b_3\rangle, \\
\beta &= \beta_* + b_3 - \langle b_3\rangle.
\end{aligned}
$$

(3.26)

The terms $b_1$, $b_2$, and $b_3$ are corrections to the Galerkin approximation. They are given, again suppressing dependence on $\phi$ and $\varepsilon$ in $W$, by

$$(3.27) \qquad b_1 = (I - P_N)\Big\{F(\tilde{u} + \varepsilon W_{k+1}^1(\tilde{u})) - F(\tilde{u}) + G(\tilde{u} + \varepsilon W_{k+1}^1(\tilde{u}), \phi; \varepsilon)\Big\},$$

$$b_2 = \big[I + \varepsilon \partial_{\tilde{u}} W_{k+1}^1(\tilde{u})\big]^{-1}$$

$$P_N\Big\{F(\tilde{u} + \varepsilon W_{k+1}^1(\tilde{u})) - F(P_N\tilde{u} + \varepsilon W_{k+1}^1(\tilde{u}))$$

$$(3.28) \qquad + G(\tilde{u} + \varepsilon W_{k+1}^1(\tilde{u}), \phi; \varepsilon) - G(P_N\tilde{u} + \varepsilon W_{k+1}^1(\tilde{u}), \phi; \varepsilon)\Big\},$$

$$b_3 = \big[I + \varepsilon \partial_\phi W_{k+1}^2(\tilde{u})\big]^{-1}$$

$$\Big\{\frac{1}{\varepsilon}\{\Omega(\tilde{u} + \varepsilon W_{k+1}^1(\tilde{u})) - \Omega(P_N\tilde{u} + \varepsilon W_{k+1}^1(\tilde{u}))\}$$

$$(3.29) \qquad + g(\tilde{u} + \varepsilon W_{k+1}^1(\tilde{u}), \phi; \varepsilon) - g(P_N\tilde{u} + \varepsilon W_{k+1}^1(\tilde{u}), \phi; \varepsilon)\Big\}.$$

All three terms converge to zero when $N \to \infty$. Exploiting Gevrey regularity of the variable $u$, we can even find exponential smallness. The main observation is, that for any $u$ with $|u|_{\mathcal{G}_{\sigma,p}} \leq R$, we have

$$|u - P_N u|_X \leq C \exp(-c_3 N^p)$$

from Hypothesis 2.4. Expanding the $b_j$ in $u$, for fixed $\psi$, this estimate immediately leads to

$$(3.30) \qquad\qquad\qquad |b_1|_X \quad \leq \quad C \exp(-c_3 N^p),$$
$$(3.31) \qquad\qquad\qquad |b_2|_X \quad \leq \quad C \exp(-c_3 N^p),$$
$$(3.32) \qquad\qquad\qquad |b_3| \quad \leq \quad C \exp(-c_3 N^p),$$

for some positive, $N$- and $\varepsilon$-independent constants $C$ and $c_3$. With the choice of $N$ from (3.2), all of the $b_j$ are bounded by

$$|b_1|_X + |b_2|_X + |b_3| \leq C' \exp\left(-c'\varepsilon^{-\frac{p}{1+\rho}}\right).$$

Step 1, (3.22) and (3.23) give an additional exponentially small contribution to the non-adiabatic terms, with exponent $-\rho/(1 + \rho)$. The estimate for the $b_j$ is improving with smaller $\rho$, the estimate for $\alpha_*$ and $\beta_*$ is improving with larger $\rho$. The optimal choice is to have the exponent equal in both cases, which is achieved by setting

$$(3.33) \qquad\qquad\qquad\qquad \rho := p.$$

Summarizing, this choice of $\rho$ in (3.33) leads to the estimate for the non-adiabatic terms from Theorem 1. For all $\tilde{u}$ with $|\tilde{u}|_{\mathcal{G}_{\sigma,p}} < R$, we have

$$|\alpha(\tilde{u}, \psi; \varepsilon)|_X \quad \leq \quad C \exp\left(-c\varepsilon^{-\frac{p}{1+p}}\right),$$

$$|\beta(\tilde{u}, \phi; \varepsilon)| \quad \leq \quad C \exp\left(-c\varepsilon^{-\frac{p}{1+p}}\right).$$

It remains to verify the claims on boundedness of the transformed nonlinearities in Theorem 1. We already observed in Step 1 that $\alpha_*$, $\tilde{F}_*$, $\beta_*$, and $\Omega_*$ are bounded, with $\tilde{F}_*$ even $O(\varepsilon^{\frac{p}{p+1}})$. It is therefore sufficient to show, that $b_1$, $b_2$, and $b_3$ are bounded in $X$ for bounded values of $\tilde{u}$ in $X$. Since $W, F$ and $G$ are bounded, $b_1$ and $b_2$ are bounded as well. The last correction, $b_3$, is bounded by $C/\varepsilon$, which gives boundedness of $\tilde{\Omega}$ in $X$. A more careful inspection of $b_3 - \langle b_3 \rangle$ shows that, indeed,

$\beta$ is bounded as well. The claims of smallness of $\tilde{F}$ and $\tilde{\Omega}$ as functions from $\mathcal{G}_{\sigma,p}$ to $\mathcal{G}_{\sigma,p}$ and $\mathbb{R}$, respectively, follow from those for $\tilde{F}$ and $b_3$. This concludes the proof of Theorem 1.                                                                                   $\square$

*Remark* 3. We briefly comment on proofs of some of the extensions mentioned in Section 2.3.

Equivariance of $A, F, G, \Omega, g$ with respect to a bounded representation of a finite-dimensional Lie group $\Gamma$ is preserved under the averaging procedure. We have to assume that $\Gamma$ leaves the Fourier approximation spaces $P_N X$ invariant. Then, the $W_k$ given in (3.6) are equivariant and the resulting, transformed equation (2.6) is $\Gamma$-equivariant also.

If $A$ is a bounded operator, we do not need Step 2 and can bound $|Au|$ without approximation in (3.20). This will lead to an exponent $-c\varepsilon^{-1}$.

The construction presented here is discontinuous in $\varepsilon$, because the number of averaging steps and the Galerkin approximation are increased discretely. However, we can smoothly interpolate in $\varepsilon$ to obtain transformations and resulting equations smoothly depending on $\varepsilon$.

3.2. **Proof of Theorem 2.** We adapt the strategy from the proof of Theorem 1 to the case of Hamiltonian systems. The additional difficulty lies solely in assuring that the transformations preserve the symplectic structure—while respecting the estimates from the iterative procedure, (3.10)–(3.13).

Again, the non-adiabatic terms derive from two sources, first the approximation of the true system by a Galerkin projection, then the non-convergence of the averaging procedure in the approximation.

Approximation with the Galerkin projection is achieved by replacing the Hamiltonian $H$ by $H_N = H \circ (Q_N, Q_N, \mathrm{id}, \mathrm{id})$, with projections $Q_N$ from Hypothesis 2.7.

The associated Hamiltonian system becomes an abstract ordinary differential equation with vector field bounded on bounded subsets of $Z \times Z \times \mathbb{R} \times S^1$ or the corresponding Gevrey class. The vector field does not depend on $(\mathrm{id} - Q_N)u$ and $(\mathrm{id} - Q_N)v$ and leaves the subset $(\mathrm{id} - Q_N)u_N = 0; (\mathrm{id} - Q_N)v_N = 0$ invariant—which follows directly from the definition of the symplectic structure and orthogonality of the $Q_N$.

The difference $H - H_N$ is small on Gevrey classes, arguing as in the proof of Theorem 1. We therefore concentrate on the bounded part of the Hamiltonian $H_N$ and drop the index $N$ for the sake of notation. Since the vector field does not depend on the "small-scale" part in $\mathrm{Ker}\,(Q_N)$, it is sufficient to construct the transformation in $Z_N := \mathrm{Rg}\,(Q_N)$.

The most common way to construct symplectic, near-identity coordinate changes is as time-one maps of Hamiltonian vector fields; see, for example, [LoMe88]. We adopt this strategy and construct the transformation

$$\Phi(\tilde{u}, \tilde{v}, \phi, J) = \mathrm{id} + \varepsilon W(\tilde{u}, \tilde{v}, \phi, J)$$

as the time-1 map of an abstract Hamiltonian differential equation on $Z_N \times Z_N \times \mathbb{R} \times S^1$, with Hamiltonian $\varepsilon\chi$.

After $k$ averaging steps, we find a general Hamiltonian of the form

$$H_k(u, v, \psi, I) = \frac{1}{2}\left((Lu, u) + (Lv, v)\right) + \frac{\Omega(I)}{\varepsilon} + \tilde{H}_{0,k}(u, v, I) + \tilde{H}_{1,k}(u, v, I, \psi),$$

where we suppressed the dependence on $\varepsilon$. The symplectic transformation $\Phi_{k+1}$ is now constructed such that the absolute value of the non-adiabatic part $\tilde{H}_{1,k+1}$ for

Gevrey data is decreased by a factor 2 compared to $\tilde{H}_{1,k}$. This task is accomplished as follows. We may assume that the average over the fast phase $\psi$ vanishes, $\langle \tilde{H}_{1,k} \rangle = 0$. We next expand $\tilde{H}_{1,k}$ in a Fourier series

$$(3.34) \qquad \tilde{H}_{1,k}(u,v,\psi,I) = \sum_{0 \neq j \in \mathbb{Z}} \tilde{H}_{1,k}^{(j)}(u,v,I) \mathrm{e}^{2\pi \mathrm{i} j \psi}.$$

The Hamiltonian $\varepsilon \chi_{k+1}$ is now explicitly defined through
$$(3.35)$$
$$\varepsilon \chi_{k+1}(\tilde{u},\tilde{v},\phi,J) = -\varepsilon \left( \Omega(J) + \varepsilon \tilde{H}_{0,k}(u,v,J) \right)^{-1} \sum_{0 \neq j \in \mathbb{Z}} \tilde{H}_{1,k}^{(j)}(\tilde{u},\tilde{v},J) \frac{1}{2\pi \mathrm{i} j} \mathrm{e}^{2\pi \mathrm{i} j \psi}.$$

This change of coordinates removes the leading-order terms as we will show now by means of a Lie series expansion; see again [LoMe88] for some background on Lie series expansions. Denote by $\{,\}$ the Poisson bracket according to the symplectic structure and by $\chi_{k+1}^{(j)}$ the $j$th Fourier coefficient of $\chi_{k+1}$. In our case, the Poisson bracket of two functions $F_1, F_2 : Z \times Z \times S^1 \times \mathbb{R} \to \mathbb{R}$ is given by

$$\{F_1, F_2\}(u,v,\psi,I) = (\partial_u F_1, \partial_v F_2)_Z - (\partial_v F_1, \partial_u F_2)_Z + \partial_\psi F_1 \cdot \partial_I F_2 - \partial_I F_1 \cdot \partial_\psi F_2.$$

We find the formal expansion

$$
\begin{aligned}
H_{k+1}(\tilde{u},\tilde{v},\phi,J) &= H_k \circ \Phi_{k+1}(\tilde{u},\tilde{v},\phi,J) \\
&= \frac{1}{2} \left( (Lu,u) + (Lv,v) \right) + \tilde{H}_{0,k}(\tilde{u},\tilde{v},J) + \frac{\Omega(J)}{\varepsilon} \\
&\quad + \tilde{H}_{1,k}(\tilde{u},\tilde{v},J,\phi) + \{H_k, \varepsilon\chi_{k+1}\} + \text{h.o.t.} \\
&= \frac{1}{2} \left( (Lu,u) + (Lv,v) \right) + \tilde{H}_{0,k}(\tilde{u},\tilde{v},J) + \frac{\Omega(J)}{\varepsilon} \\
&\quad + \tilde{H}_{1,k}(\tilde{u},\tilde{v},J,\phi) \\
&\quad - \sum_{j \in \mathbb{Z}} \frac{1}{\varepsilon} \partial_J \left( \Omega(J) + \tilde{H}_{0,k}(\tilde{u},\tilde{v},J) \right) 2\pi \mathrm{i} j \varepsilon \chi_{k+1}^{(j)} \mathrm{e}^{2\pi \mathrm{i} j \phi} + \text{h.o.t.} \\
&= \frac{1}{2} \left( (Lu,u) + (Lv,v) \right) + \tilde{H}_{0,k}(\tilde{u},\tilde{v},J) + \frac{\Omega(J)}{\varepsilon} + \tilde{H}_{1,k}(\tilde{u},\tilde{v},J,\phi) \\
&\quad - \sum_{j \in \mathbb{Z}} \partial_J(\Omega(J) + \varepsilon \tilde{H}_{0,k}(\tilde{u},\tilde{v},J)) \frac{\tilde{H}_{1,k}^{(j)}(\tilde{u},\tilde{v},J)}{\partial_J(\Omega(J) + \varepsilon \tilde{H}_{0,k}(\tilde{u},\tilde{v},J))} \mathrm{e}^{2\pi \mathrm{i} j \phi} \\
&\quad + \text{h.o.t.}
\end{aligned}
$$

Indeed, the transformation removes formally the lowest-order term depending on $\phi$. In order to estimate the transformation, later, we express $\varepsilon W_{k+1}$ in terms of the vector field, given by the partial derivatives of the Hamiltonian $\varepsilon \chi_{k+1}$,

$$(3.36) \quad \varepsilon W(\tilde{u},\tilde{v},J,\phi) = \varepsilon \begin{pmatrix} W_{\tilde{u}} \\ W_{\tilde{v}} \\ W_J \\ W_\phi \end{pmatrix} = \varepsilon \int_0^1 \begin{pmatrix} \partial_{\tilde{v}}\chi(\tilde{u}(\tau),\tilde{v}(\tau),J(\tau),\phi(\tau)) \\ -\partial_{\tilde{u}}\chi(\tilde{u}(\tau),\tilde{v}(\tau),J(\tau),\phi(\tau)) \\ \partial_\phi \chi(\tilde{u}(\tau),\tilde{v}(\tau),J(\tau),\phi(\tau)) \\ -\partial_I \chi(\tilde{u}(\tau),\tilde{v}(\tau),J(\tau),\phi(\tau)) \end{pmatrix} \mathrm{d}\tau;$$

see also [LoMe88], Appendix 7, Lemma 1. For convenience, we omitted the indices $k + 1$ of the functions $W$ and $\chi$. In the remainder of this section, we show how to obtain quantitative estimates for the higher-order terms. We write the transformed

Hamiltonian in the form

$$H_{k+1}(u, v, \phi, J) = H_k \circ \Phi$$
$$= \frac{1}{2}\left((Lu, u) + (Lv, v)\right) + \frac{\Omega(J)}{\varepsilon} + \tilde{H}_{0,k}(u, v, J) + \tilde{H}_{1,k}(u, v, J) + h_k(u, v, J, \phi).$$

We set

$$\tilde{H}_{0,k+1}(u, v, J) := \tilde{H}_{0,k}(u, v, J) + \langle h_k(u, v, J, \cdot)\rangle,$$
$$\tilde{H}_{1,k+1}(u, v, J) := \tilde{H}_{1,k}(u, v, J) + h_k(u, v, J, \psi) - \langle h_k(u, v, J, \cdot)\rangle.$$

We shall inductively show that

$$(3.37) \qquad\qquad |\tilde{H}_{1,k}|_{\mathcal{Y}+\delta_k} \leq M_k = 2^{-k}M,$$
$$(3.38) \qquad\qquad \|D\tilde{H}_{1,k}\|_{\mathcal{Y}+\delta_k} \leq M_k,$$
$$(3.39) \qquad\qquad |\tilde{H}_{1,k}|_{\mathcal{Y}+\delta_k} \leq M,$$
$$(3.40) \qquad\qquad \|D\tilde{H}_{1,k}\|_{\mathcal{Y}+\delta_k} \leq M,$$

where $\mathcal{Y} = Z_N \times Z_N \times \mathbb{R} \times S^1$, with the Gevrey norm, and $\mathcal{Y}+\delta_k$ is the extension in a complex strip, and $D$ denotes the Fréchet derivative. The sequence $\delta_k = \delta - k\eta(\varepsilon)$ is defined as in the previous section, (3.14).

We have to estimate $h_k$ in each step. Using the Cauchy estimate, Lemma 3.1, and the explicit formula for $\varepsilon W_{k+1}$, (3.36), we can estimate

$$(3.41) \qquad\qquad |\varepsilon W_{k+1}|_{\mathcal{Y}+\delta_{k+1}} \leq \frac{M_k \varepsilon}{\eta(\varepsilon)}.$$

This gives us an estimate for the correction to the Hamiltonian $h_{k+1}$,

$$|h_{k+1}|_{\mathcal{Y}+\delta_k} = |\tfrac{1}{2}\left(\varepsilon(LW_{\tilde{u}}, \tilde{u} + \varepsilon W_{\tilde{u}}) + \varepsilon(L\tilde{u}, W_{\tilde{u}}) + \varepsilon(LW_{\tilde{v}}, \tilde{v} + \varepsilon W_{\tilde{v}}) + \varepsilon(L\tilde{v}, W_{\tilde{v}})\right)$$

$$+ \frac{\Omega(J + \varepsilon W_J)}{\varepsilon} + \tilde{H}_{0,k}(\tilde{u} + \varepsilon W_{\tilde{u}}, \tilde{v} + \varepsilon W_{\tilde{v}}, J + \varepsilon W_J)$$

$$+ \tilde{H}_{1,k}(\tilde{u} + \varepsilon W_{\tilde{u}}, \tilde{v} + \varepsilon W_{\tilde{v}}, J + \varepsilon W_J, x\phi + \varepsilon W_\phi)$$

$$- \tilde{H}_{0,k}(\tilde{u}, \tilde{v}, J) - \frac{\Omega(J)}{\varepsilon} - \tilde{H}_{1,k}(\tilde{u}, \tilde{v}, J, \psi)|_{\mathcal{Y}+\delta_k}$$

$$\leq 4M_k N\varepsilon + |\Omega''|_{\mathcal{Y}+\delta_k}\varepsilon M_k + \varepsilon(M + M_k)M_k.$$

Then

$$\|h'_{k+1}\|_{\mathcal{Y}+\delta_{k+1}} \leq \varepsilon M_k \frac{4N + |\Omega''|_{\mathcal{Y}+\delta_k} + M + M_k}{\eta(\varepsilon)}.$$

For the same choice of $\eta(\varepsilon) = \eta_0 \varepsilon^{\frac{p}{1+p}}$, both $h$ and $h'$ can therefore be estimated by $\frac{M_k}{4}$, provided $\varepsilon$ is sufficiently small and $\eta_0$ sufficiently large. Proceeding with the induction as in the proof of the non-Hamiltonian version, we can perform

$$K(\varepsilon) \leq \frac{\delta}{\eta_0}\varepsilon^{-\frac{p}{1+p}}$$

transformations. Choosing $K(\varepsilon)$ maximal, the non-adiabatic dependence on the fast phase becomes exponentially small in $\varepsilon$,

$$|\tilde{H}_{1,K(\varepsilon)}|_{\mathcal{Y}+\delta/2} \leq C\exp(-c_2\varepsilon^{-\frac{p}{1+p}})$$

for $(u, v)$ bounded in $\mathcal{G}_{\sigma,p}$. Furthermore, $\tilde{H}_{1,K(\varepsilon)}$ and $\tilde{H}_{1,K(\varepsilon)}$ are bounded for $(u, v)$ bounded in $X$. Proceeding as in the proof of Theorem 1, the terms in the complement of $\mathrm{Rg}(P_N)$ can be estimated again using Gevrey regularity. Note that

$A$ is equivalent to $(\mathrm{i}L, -\mathrm{i}L)$, and therefore $|A|$ is conjugate to $(|L|, |L|)$, a selfadjoint operator. The exponential decay properties from Hypothesis 2.4 therefore follow from the resolvent estimate in 2.7, (iv). This concludes the proof of Theorem 2.    □

## 4. EXAMPLE: NONLINEAR SCHRÖDINGER EQUATIONS

We consider a system of nonlinear Schrödinger equations on the entire space $x \in \mathbb{R}^n$, or with periodic boundary conditions, alternatively. We couple the partial differential equations to a fast, anharmonic oscillator. More precisely, consider

$$(4.1) \qquad \mathrm{i}\partial_t \mathcal{A} = \Delta \mathcal{A} + \partial_{\bar{\mathcal{A}}} V(\mathcal{A}, \bar{\mathcal{A}}),$$

where $\mathcal{A} \in \mathbb{C}^m$, $\bar{\mathcal{A}}$ is the complex conjugate, and the potential $V$ is analytic in both arguments and invariant under phase rotation

$$V(\mathrm{e}^{\mathrm{i}\varphi}\mathcal{A}, \mathrm{e}^{-\mathrm{i}\varphi}\bar{\mathcal{A}}) = V(\mathcal{A}, \bar{\mathcal{A}}) \in \mathbb{R}.$$

We take $x \in \mathbb{R}^n$ or $x \in [0, \ell_1] \times \ldots \times [0, \ell_n]$ with periodic boundary conditions and consider (4.1) on the phase space $A \in X = L^2$, with values in $\mathbb{C}^m \simeq \mathbb{R}^m \times \mathbb{R}^m$.

In addition to this partial differential equation, we consider a fast one-degree-of-freedom oscillator, which we write in action-angle coordinates $I \in \mathbb{R}$, $\psi \in S^1$,

$$\frac{\mathrm{d}I}{\mathrm{d}t} = 0, \qquad \frac{\mathrm{d}\psi}{\mathrm{d}t} = \frac{\omega(I)}{\varepsilon}.$$

The scaled frequency $\omega(I) = \Omega'(I)$ is supposed to be strictly positive, uniformly in $I \in \mathbb{R}$.

Translating into the formalism of Section 2.2, we write $\mathcal{A} = u + \mathrm{i}v$ and define an unperturbed Hamiltonian function on $\mathcal{X} = X \times \mathbb{R} \times S^1$ by

$$
\begin{aligned}
H^0(u, v, I) \quad = \quad & \frac{1}{2}\left((\nabla u, \nabla u)_{L^2} + (\nabla v, \nabla v)_{L^2}\right) \\
& + \int_{\mathbb{R}^n} V(u(x) + \mathrm{i}v(x), u(x) - \mathrm{i}v(x))\mathrm{d}x + \frac{\Omega(I)}{\varepsilon}.
\end{aligned}
$$

The symplectic form on $X$ is defined via the $L^2$-scalar product, as explained in Section 2.2.

There are now various possible choices for coupling the partial differential equation to the anharmonic oscillator. A simplest example is given by coupling the angle $\psi$ to the first component,

$$(4.2) \qquad H_1(u, v, I, \psi) = I \int_{\mathbb{R}^n} (u_1 \cos(2\pi\psi) - v_1 \sin(2\pi\psi))\, \mathrm{d}x,$$

which preserves the gauge symmetry $\mathcal{A} \mapsto \mathrm{e}^{\mathrm{i}\varphi}\mathcal{A}$, $\psi \mapsto \psi - \varphi$. In complex notation, $H_1$ is simply given by $\int \mathrm{Re}\,(\mathcal{A}_1 I \mathrm{e}^{\mathrm{i}\psi})$.

We might as well break this symmetry, using a coupling of the form

$$(4.3) \qquad H_2(u, v, I, \psi) = \frac{1}{2}I \int_{\mathbb{R}^n} \left((u_1^2 - v_1^2) \cos(2\pi\psi) - 2uv \sin(2\pi\psi)\right) \mathrm{d}x,$$

which corresponds to $\frac{1}{2}\int_{\mathbb{R}^n} \mathrm{Re}(\mathcal{A}_1^2 I \mathrm{e}^{\mathrm{i}\psi})dx$.

Adding these two non-adiabatic coupling terms, we have to discuss Hamiltonian functions given by

$$H(u, v, I, \psi) = H^0(u, v, I) + \gamma_1 H_1(u, v, I, \psi) + \gamma_2 H_2(u, v, I, \psi), \qquad \gamma_1, \gamma_2 \in \mathbb{R}.$$

If we write $\mathcal{B} = I\mathrm{e}^{-\mathrm{i}\psi}$, and specialize to $m = 1$, the coupled system of evolution equations can be written in the compact form

$$\mathrm{i}\partial_t \mathcal{A} = \Delta\mathcal{A} + \partial_{\bar{\mathcal{A}}} V(\mathcal{A}, \bar{\mathcal{A}}) + \gamma_1 \mathcal{B} + \gamma_2 \bar{\mathcal{A}}\mathcal{B},$$
$$\mathrm{i}\partial_t \mathcal{B} = \frac{\omega(|\mathcal{B}|)}{\varepsilon}\mathcal{B} + \gamma_1 \int \mathcal{A} + \gamma_2 \int \mathcal{A}^2.$$

We claim that Theorem 2 can be applied to the Hamiltonian system (4.3). We therefore have to check Hypotheses 2.7 to 2.10.

Galerkin approximations, Hypothesis 2.7, are easily constructed by a cut-off in Fourier space. More precisely, we define

$$(\hat{P}_m \hat{A})(k) = \theta_m(|k|)\hat{A}(k),$$

with $\theta_m(|k|) = 1$ for $|k|^2 \leq m$ and $\theta_m(|k|) = 0$, otherwise. The required properties of $P_m$ then follow easily. The assumption on the frequency, Hypothesis 2.8, only requires $\Omega'$ to be strictly positive. The assumption on zero mean, Hypothesis 2.9, is easily checked. The assumptions on analyticity of the nonlinearities are satisfied in the interpolation space $X^s = H^{2s}$ if $2 > 2s > n/2$, that is, for the (realistic) case of $n \leq 3$; see the remark in Section 2.3. The analyticity for the Gevrey spaces can then be checked for an also modified Gevrey space $\mathcal{G}^s_{\sigma,p} = D(|A|^s \exp(\sigma|A|^p))$ with the differential operator $A = -\mathrm{i}(\Delta - \mathrm{id})$. For $p = 1/2$ and $s > n/4$, this follows as in [FeTi98], [Mat01] using that this Gevrey space is an algebra under pointwise multiplication. Theorem 2 now predicts that the influence of the rapid forcing can be adiabatically eliminated up to terms which are extremely, exponentially, small in the forcing frequency $\omega/\varepsilon$.

## 5. LOWER ESTIMATES

In this section, we give an example which shows that the exponential estimate for the non-adiabatic effects with $\varepsilon$-exponent $-p/(1+p)$ in the exponential estimate (2.7) cannot be improved to an estimate $\exp(-c/\varepsilon)$, which would hold for ordinary differential equations. Partial differential equations are genuinely different from ordinary differential equations—as far as averaging is concerned!

More precisely, our example shows that for a large class of "averaging-type" transformations, there exists a lower estimate with an exponent $-c\varepsilon^{-p}$ on the non-adiabatic terms.

Throughout, we consider a simplified version of (2.1),

$$(5.1) \qquad \frac{\mathrm{d}}{\mathrm{d}t}u = Au + F(u) + G(u, \frac{t}{\varepsilon}; \varepsilon).$$

Averaging transformations like the ones we constructed in the proofs of the main theorems are of the form $\mathrm{id} + \varepsilon W(., \frac{t}{\varepsilon}; \varepsilon)$. The transformed equation can then be written in the general form

$$(5.2) \qquad \frac{\mathrm{d}}{\mathrm{d}t}u = Au + F(u) + \tilde{F}(u, \varepsilon) + \alpha(u, \frac{t}{\varepsilon}; \varepsilon).$$

**Definition 5.1.** A transformation $\mathrm{id} + \varepsilon W(., \frac{t}{\varepsilon}; \varepsilon)$ is called a *transformation of averaging-type*, if

- (i) the transformation is close to the identity both in the $X$- and $\mathcal{G}_{\sigma,p}$-topology; more precisely, we require $W(., \tau; \varepsilon)$ to be bounded as a function from $B_X(R)$ to $X$ and as a function from $B_{\mathcal{G}_{\sigma,p}}(R)$ to $\mathcal{G}_{\sigma,p}$;

(ii) the function $W(u, \tau; \varepsilon)$ is 1-periodic in $\tau$, and we fix $W(u, 0; \varepsilon) = 0$;

(iii) if $F$ and $G$ are linear functions of $u$, then $W$ and the transformed equation are linear as well;

(iv) autonomous subspaces are preserved; more precisely, if there exists a linear projection $P$ with

$$AP = PA, \qquad PF = PF \circ P, \qquad PG \equiv 0,$$

then

$$PW \equiv 0, \qquad P\alpha \equiv 0.$$

Proceeding by induction, it is not difficult to check that the transformation given in Theorem 1 is of averaging-type.

For the rest of this section, we specialize to a *linear* Schrödinger equation with a rapidly oscillating, nonlocal potential influence. Consider

$$(5.3) \qquad \partial_t u = \mathrm{i} u_{xx} + \mathrm{i} \exp\left(2\pi \mathrm{i} \frac{t}{\varepsilon}\right) \sum_{k \in \mathbb{Z}, k \neq 0} u^k,$$

with initial value $u(0) = u_0 \in H^1(S^1, \mathbb{C})$. Here, the $u^k$ denote the spatial Fourier coefficients of the $2\pi$-periodic function $u(t, \cdot)$, $u = \sum_{k \in \mathbb{Z}} u^k \exp(\mathrm{i}kx)$. This equation fits into the framework of Theorem 1 with

$$Au = \mathrm{i}\partial_{xx} u, \quad F(u) = 0, \quad G\left(u, \frac{t}{\varepsilon}; \varepsilon\right) = \mathrm{i} \exp\left(2\pi \mathrm{i} \frac{t}{\varepsilon}\right) \sum_{k \in \mathbb{Z}, k \neq 0} u^k.$$

We consider this abstract linear evolution equation on the function space $X = H^1(S^1, \mathbb{C})$, or the corresponding Gevrey classes $\mathcal{G}_{\sigma, p}$. Since the forcing $G$ is linear and bounded in $u$, it is analytic when considered as an operator on either $X$ or $\mathcal{G}_{\sigma, p}$. It is not difficult to see that for $p \geq \frac{1}{2}$ and $\sigma > 0$, the functions in $\mathcal{G}_{\sigma, p}$ are real analytic in the spatial coordinate $x$.

**Proposition 5.2.** *Fix a function space $\mathcal{G}_{\sigma, p}$. Then for all transformations of averaging-type* $\mathrm{id} + \varepsilon W(., \frac{t}{\varepsilon}; \varepsilon)$ *of* (5.3), *there exists a sequence* $\varepsilon_k \to 0$ *such that the non-autonomous term in the transformed equation is estimated from below by*

$$(5.4) \qquad \sup_{|u|_{\mathcal{G}_{\sigma, p}} < R, \tau \in (0, 1)} |\alpha(u, \tau; \varepsilon_k)|_X \geq \frac{R}{\sqrt{2\pi}} \varepsilon_k^{1/2} \exp\left(-\sigma(2\pi)^p \varepsilon_k^{-p}\right).$$

*Note that the estimate does not depend on the choice of the transformation $W$.*

*Proof.* We exploit the definition of transformations of averaging-type, Definition 5.1. The transformation $\mathrm{id} + \varepsilon W(., \frac{t}{\varepsilon}; \varepsilon)$ is linear since the original equation is; see (iii). If we project the dynamics on the subspace of functions with mean value zero, $u^0 = 0$, then the dynamics are autonomous. The associated projection in Definition 5.1, (iv) is

$$P\left(\sum_{k \in \mathbb{Z}} u^k \exp(\mathrm{i}kx)\right) = \sum_{k \in \mathbb{Z}, k \neq 0} u^k.$$

In particular, Fourier modes with index $k \neq 0$ remain unchanged under the transformation. Hence, the transformation is of the form

$$u = v + \varepsilon W\left((v^k)_{k \in \mathbb{Z}}, \frac{t}{\varepsilon}\right),$$

with $PW \equiv 0$ and $W$ linear in $v$. This gives

$$W\left((v^k)_{k\in\mathbb{Z}}, \tau\right) = \sum_{k\in\mathbb{Z}} w^k\left(\tau\right) v^k.$$

Writing the transformed equation in spatial Fourier modes yields

$$\frac{\mathrm{d}}{\mathrm{d}t}v^k = -\mathrm{i}k^2 v^k, \quad \text{for } k \in \mathbb{Z}, k \neq 0$$

$$\frac{\mathrm{d}}{\mathrm{d}t}v^0 = \left(1 + \varepsilon w^0\left(\frac{t}{\varepsilon}\right)\right)^{-1}$$

$$\times \left(\exp\left(2\pi\mathrm{i}\frac{t}{\varepsilon}\right) \sum_{k\in\mathbb{Z}, k\neq 0} v^k - \varepsilon \sum_{k\in\mathbb{Z}} w^k\left(\frac{t}{\varepsilon}\right)\frac{\mathrm{d}}{\mathrm{d}t}v^k - \sum_{k\in\mathbb{Z}} \frac{\mathrm{d}}{\mathrm{d}\tau}w^k\left(\frac{t}{\varepsilon}\right) v^k\right).$$

Next, consider a continous family of possible initial conditions

$$u(x, \tau, \varepsilon) = (1 + \varepsilon w^0\left(\tau, \varepsilon\right))\frac{R}{|\exp(\mathrm{i}kx)|_{G_{\sigma,p}}} \exp(\mathrm{i}kx).$$

Because of (ii), we have

$$u(., 0, \varepsilon) = u(., 1, \varepsilon) = \frac{R}{|\exp(\mathrm{i}kx)|_{G_{\sigma,p}}} \exp(\mathrm{i}kx) = \bar{v}_k,$$

for all $\varepsilon$. We solve the differential equation for $v^k$, insert the result into the equation for $v^0$, and then the terms involving $w^0$ cancel. We expand the right-hand side of this equation in a Fourier series with respect to the fast time variable $\tau = t/\varepsilon$. Then the Fourier coefficient for $\exp\left(2\pi\mathrm{i}\tau\right)$ is

$$\bar{v}^k + (\varepsilon\mathrm{i}k^2 - 2\pi\mathrm{i})2\int_0^1 w^k(\tau, \varepsilon(\tau)) \exp(2\pi\mathrm{i}\tau)d\tau\bar{v}^k.$$

Thus for $\varepsilon_k = 2\pi/k^2$ the term $\exp\left(2\pi\mathrm{i}\frac{t}{\varepsilon}\right)\bar{v}^k$ cannot be eliminated by any choice of the $w^k$. The norm of this term gives a minimal contribution to the non-adiabatic remainder $\alpha$ of the size

$$\frac{R}{|\exp(\mathrm{i}kx)|_{G_{\sigma,p}}} = R|k|^{-1}\exp(-\sigma|k|^{2p}) = \frac{R}{\sqrt{2\pi}}\varepsilon_k^{1/2}\exp\left(-\sigma(2\pi)^p\varepsilon_k^{-p}\right),$$

which proves the lower estimate (5.4) and the proposition.                    $\square$

*Remark* 4. Specializing to real analytic initial data, $p = \frac{1}{2}$, the optimal exponent for exponential averaging of equation (5.3) lies between $-c\varepsilon^{-1/2}$ and $-c'\varepsilon^{-1/3}$. In particular, the exponent is different from the exponent in the case of ordinary differential equations, where we can achieve $-c\varepsilon^{-1}$.

   If we further specialize the averaging method to the procedure in the proof of Theorem 1, then we find in the case of the evolution equation (5.3) that the estimate (2.7) in Theorem 1 is sharp.

   Indeed, in the proof of Theorem 1, all spatial Galerkin modes of index $N$ with $N^{1+p}\varepsilon \geq C$ are neglected. In our example, the Galerkin modes coincide with the spatial Fourier modes and $N = k^2$. As in Proposition 5.2, the non-autonomous terms in the neglected modes can then be estimated from below by $\frac{R}{\sqrt{2\pi}}\varepsilon_k^{1/2}\exp\left(-\sigma(2\pi)^p\varepsilon_k^{-\frac{p}{p+1}}\right)$.

## References

[Bam99]     D. Bambusi, Nekhoroshev theorem for small amplitude solutions in nonlinear
            Schrödinger equations. *Math. Z.* **230** (1999), 345–387. MR **2000h:**35146

[FeTi98]    A. Ferrari and E. Titi, Gevrey regularity for nonlinear analytic parabolic equations,
            *Comm. Part. Diff. Eq.* **23** (1998), 1-16. MR **99a:**35116

[FiVi00]    B. Fiedler and M. I. Vishik, Quantitative homogenization of global attractors for
            reaction-diffusion systems with rapidly oscillating terms, MSAP preprint (2000).

[FoTe89]    C. Foias and R. Temam, Gevrey class regularity for the solution of the Navier-Stokes
            equations, *J. Funct. Anal.* **87** (1989), 359-369. MR **91a:**35135

[LoMe88]    P. Lochak and C. Meunier, *Multiphase Averaging for Classical Systems*, Applied
            Mathematical Sciences **72**, Springer-Verlag, New York (1988).

[MaRa94]    J. Marsden and T. Ratiu, *Introduction to Mechanics and Symmetry*, Texts in Applied
            Mathematics **17**, Springer-Verlag, New York (1994). MR **2000i:**70002

[Mat00]     K. Matthies, Homogenization of exponential order for elliptic systems in infinite
            cylinders, preprint (2000).

[Mat01]     K. Matthies, Time-averaging under fast periodic forcing of parabolic partial differen-
            tial equations: Exponential estimates, *J. Differential Equations* **174** (2001), 133-180.

[Nei84]     A. Neishtadt, The separation of motions in systems with rapidly rotating phase, *J.
            Appl. Math. Mech.* **48** (1984), 133-139. MR **86j:**34043

[Nek79]     N. N. Nekhorošev, An exponential estimate of the time of stability of nearly integrable
            Hamiltonian systems, (Russian) *Uspehi Mat. Nauk* **32** (1977), no. 6 (198), 5–66.

[Paz83]     A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential
            Equations*, Applied Mathematical Sciences **44**, Springer-Verlag, New York (1983).
            MR **85g:**47061

[Poe99]     J. Pöschel, On Nekhoroshev estimates for a nonlinear Schrödinger equation and a
            theorem by Bambusi. *Nonlinearity* **12** (1999), 1587–1600. MR **2001e:**37100

[Pro91]     K. Promislow, Time analyticity and Gevrey regularity for solutions of a class of
            dissipative partial differential equations, *Nonl. Analysis, Th. Meth. Appl.* **16** (1991),
            959-980. MR **92h:**35192

[SaScWu99]  B. Sandstede, A. Scheel, and C. Wulff, Bifurcations and dynamics of spiral waves. *J.
            Nonlinear Science* **9** (1999), 439–478. MR **2000h:**35020

[TBDHT96]   P. Takác, P. Bollerman, A. Doelman, A. van Harten, and E. Titi, Analyticity of essen-
            tially bounded solutions to semilinear parabolic systems and validity of the Ginzburg-
            Landau equation, *SIAM J. Math. Anal.* **27** (1996), 424–448. MR **96m:**35150

[Vis00]     M. Vishik, Attractors of evolution equations with rapidly oscillating terms, Proc.
            Internat. Conf. in Honor of R. Temam, Paris, to appear.

MATHEMATICAL INSTITUTE, UNIVERSITY OF WARWICK, COVENTRY CV4 7AL, UNITED KINGDOM
*E-mail address*: matthies@maths.warwick.ac.uk
*Current address*: FU Berlin, Institut für Mathematik I, Arnimallee 2-6, 14195 Berlin, Germany
*E-mail address*: matthies@math.fu-berlin.de

SCHOOL OF MATHEMATICS, UNIVERSITY OF MINNESOTA, 206 CHURCH ST. S.E., MINNEAPOLIS,
MINNESOTA 55455
*E-mail address*: scheel@math.umn.edu