First-Order Averaging Principles for Maps With Applications to Beam Dynamics in Particle Accelerators

H. Scott Dumas
Department of Mathematics
University of Cincinnati
Cincinnati, OH 45221–0025

James A. Ellison*
Department of Mathematics
University of New Mexico
Albuquerque, NM 87131

Mathias Vogt
Department of Mathematics
University of New Mexico
Albuquerque, NM 87131

Abstract. *For slowly evolving, discrete-time-dependent systems of difference equations (iterated maps), we believe the simplest means of demonstrating the validity of the averaging method at first order is by way of a lemma that we call Besjes' inequality. In this paper, we develop the Besjes inequality for identity maps with perturbations that are (i) at low-order resonance (periodic with short period) and (ii) far from low-order resonance in the discrete time. We use these inequalities to prove corresponding first-order averaging principles, together with a principle of adiabatic invariance on extended timescales; and we generalize and apply these mathematical results to model problems in accelerator beam dynamics, and to the Hénon map.*

Keywords: Averaging method; averaging principle; difference equations; iterated maps; small divisors; adiabatic invariance; accelerator beam dynamics; kick-rotate model; Hénon map

2000 MSC: 39A11; 37J40; 70K65; 70H11; 78A35; 70Fxx

1. Introduction

In broadest terms, the method of averaging (or "averaging principle") may be described as follows: to approximate the evolution of a system with motions occurring on both fast and slow timescales, one uses a simpler system obtained by somehow averaging over the fast motion of the original system. In the context of difference equations (or "iterated maps"), the most elementary situation to which the method applies occurs in periodic systems of the form

$$x_{n+1} = x_n + \varepsilon f(x_n, n) \tag{1.1}$$

where $x_n \in U \subset \mathbf{R}^d$, $n \in \mathbf{N}$, $\varepsilon > 0$ is a small parameter, and $f : U \times \mathbf{N} \to \mathbf{R}^d$ is a bounded, locally x-Lipschitz, discrete-time-dependent function of period p in n. Solutions of system (1.1) are approximated by solutions of the associated averaged system

$$y_{n+1} = y_n + \varepsilon \widehat{f}(y_n) \tag{1.2}$$

where the autonomous function $\widehat{f} : U \to \mathbf{R}^d$ (the *average of* f) is given by $\widehat{f}(y) = (1/p)\sum_{n=0}^{p-1} f(y, n)$. In this context the averaging principle asserts that solutions x_n of Eq. (1.1) and y_n of Eq. (1.2) that start at the same initial condition remain $O(\varepsilon)$-close on a discrete timescale of $O(1/\varepsilon)$. It is also often useful to use the continuous-time solutions of the corresponding averaged ODE

$$\frac{dy}{dt} = \varepsilon \widehat{f}(y) \tag{1.3}$$

to approximate the discrete-time solutions of Eq. (1.2) and hence also those of Eq. (1.1), so that we obtain the two approximation relations $x_n = y_n + O(\varepsilon)$ and $x_n = y(n) + O(\varepsilon)$ for $0 \le n \le O(1/\varepsilon)$ (see Theorem 1 below for a more precise statement, and note that y_n and $y(n)$ have different meanings).

Equation (1.1) is a special case of a more general problem which we also examine. Let $\nu \in \mathbf{R}$, $U \subset \mathbf{R}^d$, and $f : U \times \mathbf{R} \to \mathbf{R}^d$ be periodic with period 1 in its second argument. We then consider the system

$$x_{n+1} = x_n + \varepsilon f(x_n, n\nu)\,. \tag{1.4}$$

The analysis of this problem is similar to the analysis of the flow problem $dx/dt = \varepsilon f(x, t)$ when f is quasiperiodic in t with two base frequencies, since small divisors enter both problems in the same way.

*Corresponding author. Phone: (505) 277–4110 Fax: (505) 277–5505 E-mail: ellison@math.unm.edu

Clearly Eq. (1.4) reduces to Eq. (1.1) when $\nu = q/p$ is rational. For ν irrational, we know from Weyl's equidistribution theorem [Kö] that the average of $f(x, n\nu)$ over n exists and equals $\overline{f}(x) = \int_0^1 f(x,t)\,dt$. It is therefore natural to ask for what values of ν the solutions of Eq. (1.4) can be approximated by solutions of the two systems

$$y_{n+1} = y_n + \varepsilon \overline{f}(y_n) \tag{1.5}$$

and

$$\frac{dy}{dt} = \varepsilon \overline{f}(y)\,. \tag{1.6}$$

In this paper we present approximation theorems that are both theoretically satisfying and useful in applications. In particular, we weaken the usual small divisor conditions on ν (in which ν satisfies infinitely many "Diophantine conditions"), requiring instead only finitely many conditions at appropriately low order. These conditions exclude ν from zones centered on low-order rationals, and in this "far-from-low-order-resonance case" (where ν satisfies only "truncated Diophantine conditions" and is *not* necessarily irrational), we again find that $x_n = y_n + O(\varepsilon) = y(n) + O(\varepsilon)$ for $0 \le n \le O(1/\varepsilon)$ (see Theorem 2 below). Under the additional hypothesis that the average of the perturbation vanishes, we are able to show adiabatic invariance of solutions of system (1.4) on extended timescales up to $O(1/\varepsilon^2)$ (see Theorem 3).

Initially, we state Theorems 1, 2, and 3 under the hypothesis that the perturbation εf has compact support in its x-domain, which is assumed to be all of $\mathbf{R}^d$; this avoids *a priori* restrictions on ε and permits clear proofs. To obtain results better suited to applications, we then give propositions that extend our theorems to more general perturbations on more general domains, and also to more general Diophantine conditions in which the zones mentioned above are allowed to depend on ε; this in turn allows ν to come within $O(\varepsilon^\lambda)$ of low-order rationals, but with loss of accuracy in the approximation (see Propositions A and B below). Finally, we revisit the case of resonant ν in a way that permits us to explore $O(\varepsilon)$ neighborhoods of low-order resonances $\nu = q/p$: we set $\nu = q/p + \varepsilon a$ (where $a \in \mathbf{R}$ should be viewed as a measure of the $O(\varepsilon)$ displacement from the resonance) and rewrite Eq. (1.4) as the system

$$\begin{pmatrix} x_{n+1} \\ \tau_{n+1} \end{pmatrix} = \begin{pmatrix} x_n + \varepsilon\, f(x_n, \frac{q}{p}n + \tau_n) \\ \tau_n + \varepsilon a \end{pmatrix}. \tag{1.7}$$

This is in the form of Eq. (1.1) with x_n replaced by $(x_n, \tau_n)^{\mathrm{T}}$. Writing $\widehat{f}(x,\tau) = 1/p \sum_{n=0}^{p-1} f(x, nq/p + \tau)$, the averaged problem reduces to

$$\begin{pmatrix} y_{n+1} \\ \tau_{n+1} \end{pmatrix} = \begin{pmatrix} y_n + \varepsilon\, \widehat{f}(y_n, \tau_n) \\ \tau_n + \varepsilon a \end{pmatrix}, \tag{1.8}$$

and we recapture the relations $x_n = y_n + O(\varepsilon) = y(n) + O(\varepsilon)$ for $0 \le n \le O(1/\varepsilon)$, where $y(t)$ is the solution of the system

$$\frac{d}{dt}\begin{pmatrix} y \\ \tau \end{pmatrix} = \varepsilon \begin{pmatrix} \widehat{f}(y,\tau) \\ a \end{pmatrix}, \tag{1.9}$$

which is equivalent to the non-autonomous system $dy/dt = \varepsilon \widehat{f}(y, \varepsilon a t)$; see Proposition C below.

Using the generalized versions of our theorems (provided by Propositions A, B, and C), we obtain an essentially complete description of solutions of system (1.4) on $O(1/\varepsilon)$ timescales for various values of ν (there are however thin gaps at the boundaries between the ν for which resonant and nonresonant motions occur; cf. Remark 2.5 below).

Averaging principles for maps are not new; results in this direction have been available since the 1960s (cf. for example [Bel], [Dr]), and related normal form methods have been used extensively in accelerator beam dynamics (cf. [BTT] and references therein). However, we believe that by admitting perturbations with ν far from low-order resonance (more general than highly nonresonant perturbations), our methods will be especially useful in applications, and we give examples from accelerator beam dynamics below. We also believe that our proofs are new in an important way, as they directly compare the solutions x_n and y_n using certain simple but ingenious inequalities, the first of which appears below in Lemma 1 (§4.1). Because the first explicit version of this lemma seems to occur in [Bes], we call it the *Besjes inequality*, though it was probably inspired by more general methods for nonperiodic systems as presented in [BM], and was probably also known to Russian researchers earlier [Ne]. For the elementary system (1.1), the difference between this and other methods of proof (e.g. using a near-identity transformation) may seem slight. However, in applications and other contexts, the difference can be significant, for reasons that are partly explained in Remarks 4.1 and 4.3 below. We therefore believe that the Besjes inequality and its generalizations should be of special interest to researchers working in both theory and application of averaging methods.

The remainder of this paper is organized as follows. In Section 2 we present the details of our averaging results described informally above. In Section 3 we apply the averaging principles to model problems in accelerator beam dynamics, showing that solutions of a class of "kick-rotate" models are well-approximated by solutions of the corresponding averaged models. We also apply the adiabatic invariance principle to the Hénon map (often used to model sextupole magnets in accelerators). In Section 4, we formulate the main technical tools required to prove the results in Section 2. These are the so-called Besjes inequality for periodic functions (Lemma 1, §4.1), and its generalization to functions far from low-order resonance (Lemma 2, §4.2.2). After formulating and proving these inequalities, we use them to prove the mathematical results from Section 2. Finally, for the sake of completeness, in the Appendix we state and prove two elementary results used in earlier proofs.

We end this introduction with a few words about notation. We use the symbols $\mathbf{N}$, $\mathbf{R}$, $\mathbf{R}_+$, and $\mathbf{Z}$ to denote, respectively, the counting numbers $\{0, 1, 2, \ldots\}$, the real numbers, the positive real numbers, and the integers. The symbol $|\ |$ indicates the Euclidean norm on $\mathbf{R}^d$ (or the absolute value $|k|$ of an integer k), and $\|\ \|_S$ denotes the uniform norm of a function over the set S; i.e., $\|F\|_S := \sup_{x\in S} |F(x)|$.

2. Averaging Principles and Adiabatic Invariance

In this section we state—and provide brief remarks on—our approximation results for maps as discussed in the introduction above.

2.1 Averaging for Maps with Periodic Perturbations

Let us be more precise about the functions f in Eq. (1.1) to which our results apply. First, taking $S = \mathbf{R}^d \times \mathbf{N}$, we assume that $f : S \to \mathbf{R}$ satisfies the following:

(i) f is bounded on S and $f(\cdot, n)$ is locally Lipschitz, uniformly in n
(ii) There exists a positive integer p such that $(x, n) \in S \Rightarrow f(x, n + p) = f(x, n)$
(iii) There is an $r > 0$ such that $|x| \geq r$ and $n \in \mathbf{N} \Rightarrow f(x, n) = 0$

When f satisfies (ii), we say it is "periodic with period p in its second argument"; and when it satisfies (iii), it is "compactly supported in x, uniformly in n." It follows from (i) and (iii) that f is globally Lipschitz in x, uniformly in n. In Subsection 2.4 we show how to treat the case where f is not compactly supported.

We now state a simple averaging principle for maps with periodic perturbation $\varepsilon f(x, n)$ and corresponding averaged perturbation $\varepsilon\widehat{f}(y) = (\varepsilon/p)\sum_{n=0}^{p-1} f(y, n)$:

Theorem 1. *Let $S = \mathbf{R}^d \times \mathbf{N}$, and suppose $f : S \to \mathbf{R}^d$ satisfies assumptions (i), (ii), and (iii) above. Fix $\varepsilon \in (0, 1]$, and consider the system*

$$x_{n+1} = x_n + \varepsilon f(x_n, n) \tag{1.1}$$

together with the associated averaged systems

$$y_{n+1} = y_n + \varepsilon\widehat{f}(y_n) \qquad (1.2), \qquad\qquad \text{and} \qquad\qquad \frac{dy}{dt} = \varepsilon\widehat{f}(y)\,. \qquad (1.3)$$

Choose $T > 0$ to capture the desired properties of system (1.3) on $[0, T/\varepsilon]$. Then there exist positive constants $C = C(T)$ and $C' = C'(T)$ such that the solutions x_n, y_n, and $y(t)$ of Eqs. (1.1), (1.2), and (1.3) with common initial condition $x_0 = y_0 = y(0)$ exist uniquely for all time and satisfy $|x_n - y_n| \leq Cp\,\varepsilon$ and $|x_n - y(n)| \leq (Cp + C')\,\varepsilon$ for $0 \leq n \leq T/\varepsilon$.

2.2 Averaging for Maps With Perturbations Far From Low-Order Resonance

We now present an averaging principle for system (1.4), where ν is a fixed positive number. When we write $\nu = q/p$, we mean that q and $p > 0$ are relatively prime integers with the *order* of the rational number

ν given by $p > 0$. Using this convention, we first note that if $\nu = q/p$, then $f(x, n\nu)$ has integer period p in n, and Theorem 1 applies. In fact, as we shall see in Proposition C, Theorem 1 applies not only *at* low-order rationals but also *near* them. However, since the error estimate in this theorem is proportional to p, it is not very useful when p is "large." We therefore restrict use of Theorem 1 to situations where p is "small" (the "low-order-resonance case"), and we next focus on situations where ν is far from low-order rational numbers (the "far-from-low-order-resonance case"). In this case small divisors inevitably enter the analysis (see the proof of Lemma 2, §4.2.2) and it might be expected that ν would need to be "highly irrational" (e.g. satisfy infinitely many Diophantine conditions). We show instead that the averaging principle may be established when ν satisfies only finitely many Diophantine conditions to a certain order, and we call these *truncated Diophantine conditions.*

In more precise terms, ν satisfies truncated Diophantine conditions if it belongs to the set $\mathcal{D}(\phi, R)$ defined below in Eq. (4.3), where ϕ is the *zone function* of the Diophantine condition and $R > 0$ is the *truncation order* or *ultraviolet cutoff*, which gives precise meaning to the phrase "p large" used above (i.e., p is large if $p > R$). Roughly speaking, $\mathcal{D}(\phi, R)$ is constructed by removing open intervals centered on low-order rationals $\nu = q/p$. The zone function ϕ controls the size of the intervals removed, and the cutoff R is the maximal order of rationals from around which intervals are removed. These terms are defined precisely in Subsection 4.2.1 (to fully understand the difference between truncated and ordinary Diophantine conditions, and to appreciate the advantages offered by the former, the reader may find it worthwhile to read that subsection).

With truncated Diophantine conditions given explicitly in Eq. (4.3), we now consider the class of functions to which our next result applies. For $S = \mathbf{R}^d \times \mathbf{R}$ we consider functions $f : S \to \mathbf{R}^d$ satisfying the following conditions (analogous to (i) through (iii) in §2.1):

(j) f is of class C^4 on S

(jj) $(x, \theta) \in S \Rightarrow f(x, \theta + 1) = f(x, \theta)$

(jjj) There is an $r > 0$ such that $|x| \geq r$ and $\theta \in \mathbf{R} \Rightarrow f(x, \theta) = 0$

Terminology for describing conditions (jj) and (jjj) is similar to that for describing conditions (ii) and (iii) above in Subsection 2.1. Since we assume f has unit period in its second argument, its average $\overline{f}$ is simply $\overline{f}(y) := \int_0^1 f(y, \theta)\, d\theta$. Finally, we alert the reader that the truncated Diophantine conditions satisfied by ν must be adapted to f in the sense that the zone function ϕ must decay appropriately; this is made precise in Eq. (4.2) of Subsection 4.2.1 (basically ϕ must decay fast enough so that $\mathcal{D}(\phi, R)$ is nonempty, but slow enough so that the series in Eq. (4.2) converges; this accounts for assumption (j) above and our specific choice of ϕ as discussed in §4.2.1).

We now state our averaging principle for maps with perturbations $\varepsilon f(x, n\nu)$ far from low-order resonance and averaged perturbation $\varepsilon\overline{f}(y)$ as above:

Theorem 2. *Let $S = \mathbf{R}^d \times \mathbf{R}$, suppose $f : S \to \mathbf{R}^d$ satisfies assumptions (j), (jj), and (jjj) above, and suppose the zone function ϕ is adapted to f on $\mathbf{R}^d$ in the sense of Eq. (4.2). Fix $\varepsilon \in (0, 1]$, and consider the system*

$$x_{n+1} = x_n + \varepsilon f(x_n, n\nu) \tag{1.4}$$

together with the associated averaged systems

$$y_{n+1} = y_n + \varepsilon\overline{f}(y_n) \qquad (1.5)\,, \qquad \text{and} \qquad \frac{dy}{dt} = \varepsilon\overline{f}(y)\,. \qquad (1.6)$$

Choose $T > 0$ to capture the desired properties of system (1.6) on $[0, T/\varepsilon]$. Then there exist positive constants R_ε, $C = C(f, \phi, T)$, and $C' = C'(f, \phi, T)$ such that whenever $\nu \in \mathcal{D}(\phi, R_\varepsilon)$ (defined in Eq. (4.3)), the solutions x_n, y_n, and $y(t)$ of Eqs. (1.4), (1.5), and (1.6) with common initial condition $x_0 = y_0 = y(0)$ exist uniquely for all time and satisfy $|x_n - y_n| \leq C\,\varepsilon$ and $|x_n - y(n)| \leq C'\,\varepsilon$ for $0 \leq n \leq T/\varepsilon$.

Remark 2.1 For averaging principles of this type, it is natural to consider the average $\lim_{N\to\infty}(1/N)\sum_{n=0}^{N-1} f(x, n\nu)$ of f over n as mentioned in the introduction. Under mild integrability conditions on f, it can be shown that when ν is irrational, this average converges to $\int_0^1 f(x, \theta)\, d\theta$, which is the average used here (this is related to Weyl's equidistribution theorem; cf. [Br] and [Kö]). However, our results do not require the existence of the average of $f(x, n\nu)$ over n, nor do they require ν to be irrational; instead we require $\nu \in \mathcal{D}(\phi, R_\varepsilon)$, and this latter set contains many rationals of order greater than R_ε.

2.3 Adiabatic Invariance on Extended Timescales

In this subsection, we consider a special system somewhat like a perturbation of an integrable Hamiltonian system. As in Theorem 2, we assume that ν satisfies truncated Diophantine conditions, but now we assume additionally that the perturbation εf has zero mean; i.e., we assume that

(jw) For each $x \in \mathbf{R}^d$, $\displaystyle\int_0^1 f(x,\theta)\,d\theta = 0$

This extra hypothesis gives an averaging principle showing that the action-like variables are adiabatically invariant over timescales longer than $O(1/\varepsilon)$:

Theorem 3. *Let $S = \mathbf{R}^d \times \mathbf{R}$, suppose $f : S \to \mathbf{R}^d$ satisfies conditions (j), (jj), (jjj), and (jw) above, and suppose the zone function ϕ is adapted to f on $\mathbf{R}^d$ (as in Eq. (4.2)). Fix $\varepsilon \in (0,1]$, choose $T > 0$, and consider the system*

$$x_{n+1} = x_n + \varepsilon f(x_n, n\nu) \tag{1.4}$$

with arbitrary initial condition $x_0 \in \mathbf{R}^d$. Then there exist positive constants R_ε, $K_1 = K_1(f,\phi)$, and $K_2 = K_2(f,\phi)$ such that whenever $\nu \in \mathcal{D}(\phi, R_\varepsilon)$ (cf. Eq. (4.3)), the solution x_n of Eq. (1.4) satisfies $|x_n - x_0| \le K_1\,\varepsilon + K_2\,\varepsilon^2 n$ for $n \in \mathbf{N}$. In particular, for $0 \le \alpha \le 1$, we have $|x_n - x_0| \le C(T)\,\varepsilon^\alpha$ for $0 < n \le T/\varepsilon^{2-\alpha}$, where $C(T) = K_1 + K_2 T$.

Remark 2.2 Using second (or higher) order averaging, it is possible to get a better estimate of $|x_n - x_0|$ on the full $O(1/\varepsilon^2)$ time interval (see [ES] for a flow version).

2.4 Extensions and Generalizations

In this subsection we give three propositions that extend and generalize our results above, making them more suitable for applications. Our first proposition shows that Theorems 2 and 3 may be generalized to the case where the zones of the truncated Diophantine conditions depend on ε.

Proposition A (ε-dependent zone functions). *Suppose that $0 \le \lambda \le 1$, and that in Theorem 2 [or Theorem 3], the zone function ϕ is replaced by the new zone function $\varepsilon^\lambda\phi$. Then the conclusions of the theorem remain true, provided that the error estimates $C\,\varepsilon$ and $C'\varepsilon$ are modified to read $C\varepsilon^{1-\lambda}$ and $C'\varepsilon^{1-\lambda}$ [or $C(T)\,\varepsilon^\alpha$ is modified to read $C(T)\,\varepsilon^{\alpha-\lambda}$].*

In order to clarify and simplify the mathematical structure of our methods, we have presented Theorems 1, 2, and 3 under the assumption that the perturbations have compact support on spatial domains that are all of $\mathbf{R}^d$. Our next proposition shows that this assumption may be removed at little cost.

Proposition B (more general perturbations). *Suppose that the domain $S = \mathbf{R}^d \times \mathbf{N}$ in Theorem 1 is replaced by the more general domain $S' = U \times \mathbf{N}$, where $U \subset \mathbf{R}^d$ is open [or the domain $S = \mathbf{R}^d \times \mathbf{R}$ in Theorem 2 or 3 is replaced by $S' = U \times \mathbf{R}$, $U \subset \mathbf{R}^d$ open], and assumption (iii) is removed from the hypotheses of Theorem 1 [or (jjj) is removed from the hypotheses of Theorem 2 or 3]. Then the conclusions of Theorem 1 [or Theorem 2 or 3] remain true provided that: (a) $0 \le \varepsilon < \varepsilon_0$, where the threshold $\varepsilon_0 > 0$ may be estimated as outlined below in Subsection 4.3.2; and (b) the conclusion "exist uniquely for all time" is replaced by "exist uniquely on the time interval $[0, T/\varepsilon]$," with $T > 0$ chosen strictly less than $\beta(x_0)$, where $[0, \beta(x_0))$ is the maximal forward interval of existence for the averaged flow problem $d\breve{y}/dt' = \widehat{f}(\breve{y})$ in the domain U [or for the flow problem $d\breve{y}/dt' = \overline{f}(\breve{y})$ in U].*

Remark 2.3 Of course Proposition A also applies to Proposition B.

The following proposition shows that Theorem 1 may be used to analyze the dynamics of solutions of Eq. (1.4) in $O(\varepsilon)$ neighborhoods of low-order resonances $\nu = q/p$.

Proposition C (behavior near low-order resonance). *Let $U \subset \mathbf{R}^d$ be open, $S' = U \times \mathbf{R}$, and suppose $f : S' \to \mathbf{R}^d$ satisfies conditions (j) and (jj) of Theorem 2 with S replaced by S'. Fix the rational number q/p, $p > 0$ and q relatively prime, and fix $a \in \mathbf{R}$. Then Eq. (1.4) with $\nu = q/p + a\,\varepsilon$ may be rewritten as Eq. (1.7), and Theorem 1 together with Proposition B apply with x and y replaced by $(x,\tau)^{\mathrm{T}}$ and $(y,\tau)^{\mathrm{T}}$ respectively. In particular there are positive constants ε_0, $c = c(T,|a|)$, and $c' = c'(T,|a|)$ such that $|x_n - y_n| \le c\,p\,\varepsilon$ and $|x_n - y(n)| \le (cp + c')\,\varepsilon$ for $0 \le \varepsilon < \varepsilon_0$ and $0 \le n \le T/\varepsilon$.*

Remark 2.4 Clearly y_n evolves by $y_{n+1} = y_n + \varepsilon \widehat{f}(y_n, \varepsilon a n)$; and also $y(n) = \check{y}(\varepsilon n)$, where $\check{y}$ evolves via $d\check{y}/dt = \widehat{f}(\check{y}, at)$.

Remark 2.5 Propositions A and B characterize the motion of x_n to within $O(\varepsilon^{1-\lambda})$ for ν away from low-order rationals, i.e., outside of $O(\varepsilon^{\lambda}\phi(p)/p)$ neighborhoods of rationals q/p with $0 < p \le R_\varepsilon$. For these ν the nonresonant normal form of Eq. (1.6) applies. Proposition C characterizes the motion to within $O(\varepsilon p)$ for ν inside $O(\varepsilon)$ neighborhoods of q/p. For these ν the resonant normal form of Eq. (1.9) applies. What is missing is information about the motion for ν in the gaps between the domains of validity of the resonant normal form and the nonresonant normal form. The size of the gaps decreases to zero as $\lambda \nearrow 1$; however, the error in the nonresonant normal form simultaneously deteriorates to $O(1)$. High-order rationals, i.e. q/p with $p > R_\varepsilon$, are of course treated using Proposition B. It is interesting to note that they may also be treated using Proposition C; however, the $O(p\varepsilon + \varepsilon)$ error bound deteriorates to $O(1)$ as p approaches $O(1/\varepsilon)$.

3. Examples From Accelerator Beam Dynamics

Modern particle accelerators operate at the limits of current technology, and their design and operation depend crucially on an understanding of the dynamics of particle beams. In this section we give examples showing how Theorems 1 and 2 (supplemented by Propositions A, B and C) may be used to analyze a class of beam dynamics models, and how Theorem 3 may be used to analyze the Hénon map (which is itself a model of certain features in beam dynamics). In fact, our averaging principles for maps have features that make them especially effective for this purpose; namely, they compare solutions of the exact and averaged model problems in the simplest possible way, and produce rigorous mathematical bounds on the difference between these solutions in an essentially optimal fashion. Although $O(1/\varepsilon)$ times may be short by accelerator standards (and adiabatic invariance of actions on $O(1/\varepsilon^2)$ times is perhaps ideal), we see our work here as an important step in understanding the dynamics of maps on long timescales.

We point out that this section extends certain results of [ES] in at least two important ways: first, by using maps, we are able to incorporate delta function type "kicks" that could not be treated rigorously via the flow methods of [ES]; second, the truncated Diophantine conditions used here are more physically realistic and explicit than the small divisor conditions used there (cf. §4.2.1). We also note that our maps need not be polynomial; however, a very nice approach using normal forms for polynomial maps in beam dynamics is presented in [BTT].

We begin in Subsection 3.1 with a general "kick-rotate" model in one degree of freedom. In Subsection 3.2 we apply the results of Subsection 3.1 to the important case of the weak-strong beam-beam interaction, and in Subsection 3.3 we apply Theorem 3 and Proposition B to the Hénon map.

3.1 The One Degree of Freedom Kick-Rotate Model

In this subsection, for purposes of illustration we focus on a simple but widely used class of beam dynamics models: the so-called one degree of freedom "kick-rotate" models. We note, however, that our methods may be generalized to treat models with several degrees of freedom and at higher order (this will be the subject of a future publication [DEVS]).

A circular accelerator (in storage mode) has a closed orbit, that is, there exists a unique solution of the equations of motion which has the periodicity of the (circular) acclerator. A complete, three-degree-of-freedom description of single-particle beam dynamics involves three spatial coordinates in the co-moving (Frenet-Serret) system defined by the projection of the closed orbit on configuration space, and their three conjugate momenta. It is convenient to study the dynamics in terms of a Poincaré map (one-turn map) at a fixed azimuthal location in the ring. Here we consider one transverse degree of freedom and let w_1 and w_2 denote the spatial coordinate and conjugate momentum in the Poincaré section. The base-model consists of a "rotation with unperturbed tune ν" representing the linear "betatron motion." Perturbations of this model often consist of an instantaneous change in momentum w_2 at a fixed location in the ring, which depends only on the spatial coordinate w_1 (a "kick-map"). If we take this fixed location to be the azimuthal position of the Poincaré section, then the perturbed dynamics is given by the so-called "kick-rotate" model

$$w_{n+1} \;=\; R\,w_n \;+\; \varepsilon\,R\begin{pmatrix}0\\ -H'(w_{1,n})\end{pmatrix}, \qquad \text{where} \quad R := e^{\mathcal{J}2\pi\nu}, \tag{3.1}$$

that is, a kick followed by a rotation through the angle $2\pi\nu$ about the origin. Here $\mathcal{J} := \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}$ is the unit symplectic matrix and H' is the "kick function." Since R depends only on the fractional part of ν we shall assume $\nu \in [0,1]$ in the following. The map defined by Eq. (3.1) is symplectic since it is the composition of symplectic maps. The notation $w_{1,n}$ indicates the first component of the vector $w_n = (w_1, w_2)_n^{\mathrm{T}}$ (we hope the reader will forgive us the ambiguity of using w_n to denote a vector and w_1 or $w_{1,n}$ its first component, and w_2 or $w_{2,n}$ its second component; the meaning should be clear from context, since we rarely explicitly set $n = 1$ or $n = 2$).

For $R = 1$, i.e. $\nu \in \{0,1\}$, Eq. (3.1) is easily solved and gives $w_n = (w_{1,0}, -nH'(w_{1,0}))^{\mathrm{T}}$ and thus $|w_{2,n}|$ is monotonically increasing to infinity. For $R = -1$ (i.e., $\nu = 1/2$), $w_{2n} = (w_{1,0}, -2nH'(w_{1,0}))^{\mathrm{T}}$ and the motion is again unbounded. Thus for $\nu \in \{0, 1/2, 1\}$ and for all initial conditions where $H'(w_{1,0}) \neq 0$, the distance from the origin is monotonically increasing. The basic question is, What happens for general ν? We shall apply the results of Section 2 to answer this question for most ν in $[0,1]$.

Eq. (3.1) may be written as

$$w_{n+1} = R\,w_n + \varepsilon\,R\,F(w_n) \tag{3.2}$$

and the transformation $w_n = R^n\,x_n$ recasts Eq. (3.2) as:

$$x_{n+1} = x_n + \varepsilon\,R^{-n}\,F(R^n\,x_n) =: x_n + \varepsilon\,f(x_n, n\nu)\;, \tag{3.3}$$

which is in the standard form for averaging (cf. Eq. (1.4)).

It is easy to see that $f(x,\theta) = H'(x_1\cos 2\pi\theta + x_2 \sin 2\pi\theta)\,(\sin 2\pi\theta, -\cos 2\pi\theta)^{\mathrm{T}} = (\partial H/\partial x_2\,,\,-\partial H/\partial x_1)^{\mathrm{T}}$. Thus if we define $\mathcal{H}(x,\theta) := H(x_1\cos 2\pi\theta + x_2\sin 2\pi\theta)$, then Eq. (3.3) becomes

$$x_{n+1} = x_n + \varepsilon\,\mathcal{J}\,\nabla_x \mathcal{H}(x_n, n\nu)\;. \tag{3.4}$$

Equations (3.3) and (3.4) also define symplectic maps, since the transformation is symplectic.

3.1.1 The kick-rotate model in the far-from-low-order-resonance case

In this subsection, we examine the behavior of the kick-rotate model (3.1) in the case where the tune belongs to the ε-dependent truncated Diophantine set $\mathcal{D}(\varepsilon^\lambda\phi, R_\varepsilon)$. In physical terms, this means that the tune is "far from low-order resonance."

The most useful form of $\mathcal{H}$ in Eq. (3.4) is given in terms of the Fourier series $H(\sqrt{2J}\sin 2\pi t) = \sum_{k\in\mathbf{Z}} H_k(J)\,e^{i2\pi kt}$, from which it follows that $\mathcal{H}(x,n\nu) = \sum_{k\in\mathbf{Z}} H_k(J(x))\,e^{i2\pi k(\Phi(x)+n\nu)}$, where Φ and J are defined by $x_1 = \sqrt{2J}\sin(2\pi\Phi)$ and $x_2 = \sqrt{2J}\cos(2\pi\Phi)$. The averaged problem is then

$$y_{n+1} = y_n + \varepsilon\,\mathcal{J}\,\nabla_y H_0(J(y_n))\,, \tag{3.5}$$

where $H_0(J) = \int_0^1 H(\sqrt{2J}\sin 2\pi t)\,dt$. The associated (scaled) flow problem is

$$\frac{d\breve{y}}{dt} = 2\pi\,\omega(J(\breve{y}))\;\mathcal{J}\breve{y}\;, \qquad \breve{y}(0) = x_0 \tag{3.6}$$

where $2\pi\omega(J) = H_0'(J)$. We note that the map defined in Eq. (3.5) is only symplectic through $O(\varepsilon)$; however, the vector field in Eq. (3.6) is Hamiltonian with Hamiltonian $H(J(\breve{y}))$. It is easy to check that $J(\breve{y}) = \frac{1}{2}(\breve{y}_1^2 + \breve{y}_2^2)$ is constant along orbits so that $J(\breve{y}) = J_0 = J(x_0)$ and thus $\breve{y}(t) = e^{\mathcal{J}2\pi\omega(J_0)t}\,x_0$. Finally, Theorem 2 together with Propositions A and B give

$$w_n = e^{\mathcal{J}2\pi n(\nu+\varepsilon\omega(J_0))}\,x_0 + O(\varepsilon^{1-\lambda}) \tag{3.7}$$

for $0 \leq n \leq T/\varepsilon$, with ε suitably restricted as in Proposition B for non-compactly supported perturbations, with $\lambda \in [0,1)$, $\nu \in \mathcal{D}(\varepsilon^\lambda \phi, R_\varepsilon)$, and with R_ε defined by the condition

$$\sum_{|k|>R_\varepsilon} \|H_k'(J)\nabla_x J\|_{D(\delta)} + \|H_k(J)2\pi\nabla_x \Phi\|_{D(\delta)} \;<\; \zeta\varepsilon\,, \tag{3.8}$$

where $D(\delta)$ is the δ-tube around the solution of Eq. (3.6) (see the definition of the δ-tube in §4.3.2).

3.1.2 The kick-rotate model in the near-to-low-order-resonance case

For ν near low-order resonance, we write $\nu = \frac{q}{p} + \varepsilon a$ when p is not too large (more precisely, when $0 < p \leq R_\varepsilon$ for suitable $\zeta, \varepsilon > 0$ in (3.8)). Thus using Eq. (1.7), our problem becomes

$$\begin{pmatrix} x_{n+1} \\ \tau_{n+1} \end{pmatrix} = \begin{pmatrix} x_n + \varepsilon\,\mathcal{J}\,\nabla_x \mathcal{H}(x_n, n\frac{q}{p} + \tau_n) \\ \tau_n + \varepsilon a \end{pmatrix} \,. \tag{3.9}$$

We are now in the periodic case, with averaged Hamiltonian $\widehat{\mathcal{H}}(x,\tau) = (1/p)\sum_{n=0}^{p-1} H(x_1 \cos(2\pi[n\frac{q}{p} + \tau]) + x_2 \sin(2\pi[n\frac{q}{p} + \tau]))$. The averaged problem is $(y_{n+1}, \tau_n)^{\mathrm{T}} = (y_n + \varepsilon\,\mathcal{J}\,\nabla_y \widehat{\mathcal{H}}(y_n, \tau_n),\, \tau_n + \varepsilon a)^{\mathrm{T}}$, with its associated scaled flow $(d\check{y}/dt, d\tau/dt)^{\mathrm{T}} = (\mathcal{J}\nabla_{\check{y}} \widehat{\mathcal{H}}(\check{y}, \tau),\, a)^{\mathrm{T}}$. Solving for τ gives $\frac{d\check{y}}{dt} = \mathcal{J}\,\nabla_{\check{y}} \widehat{\mathcal{H}}(\check{y}, at)$. Theorem 1 with Propositions B and C then give

$$w_n = e^{\mathcal{J}2\pi n\nu}\, x_n = e^{\mathcal{J}2\pi n(\frac{q}{p}+\varepsilon a)}\, \check{y}(\varepsilon n) + O(\varepsilon) \tag{3.10}$$

for $0 \leq n \leq T/\varepsilon$ and for $\nu = \frac{q}{p} + \varepsilon a$. However, it is not clear we have achieved a great simplification and so we look more closely. It turns out that $\widehat{\mathcal{H}}\,(\exp(-\mathcal{J}2\pi\theta')\check{y}, \theta) = \widehat{\mathcal{H}}(\check{y}, \theta - \theta')$, which suggests that an autonomous Hamiltonian system might be found with the symplectic transformation $\check{y} \mapsto \check{z}$ defined by $\check{y} = e^{-\mathcal{J}2\pi a t}\,\check{z}$. This is indeed true and gives the autonomous system

$$\frac{d\check{z}}{dt} = 2\pi a\,\mathcal{J}\,\check{z} + \mathcal{J}\,\nabla_{\check{z}} \widehat{\mathcal{H}}(\check{z}, 0) \tag{3.11}$$

with Hamiltonian $\mathcal{K}(\check{z}) = 2\pi a J(\check{z}) + \widehat{\mathcal{H}}(\check{z}, 0)$. Equation (3.10) thus becomes

$$w_n = e^{\mathcal{J}2\pi n\frac{q}{p}}\, \check{z}(\varepsilon n) + O(\varepsilon) \tag{3.12}$$

from which the behavior of the approximation is now quite transparent.

3.1.3 Summary of the kick-rotate model

We now have the following picture of the solutions of Eq. (3.1) on $O(1/\varepsilon)$ time intervals. For $\nu \in \mathcal{D}(\varepsilon^\lambda\phi, R_\varepsilon)$ the motion is given by Eq. (3.7) and thus our kick-rotate map behaves like a twist map with tune $\nu + \varepsilon\omega(J_0)$. For these ν the effect of the perturbation is slight; the up and down kicks on the integral curves essentially cancel and the main effect of the perturbation is to create an amplitude-dependent tune. For $\nu = \frac{q}{p} + \varepsilon a$, we see that in the p-periodic Poincaré map, the approximate motion moves slowly along the phase curves given by the level curves of $\mathcal{K}(\check{z})$. We thus have an essentially complete picture of the motion (except for small gaps in ν as discussed in Remark 2.5).

3.2 The Weak-Strong Beam-Beam Effect

As a concrete example, we study the weak-strong beam-beam effect for round Gaussian beams in collider rings. We treat the lattice (the sequence of transport maps through the various components of the accelerator) as a stable, linear symplectic map, and the beam-beam interaction as localized at the point of the ring where the bunches collide (the "interaction point"). The phase space distribution of the strong beam at the interaction point is assumed to be stationary; in particular the beam-beam effect of the weak beam on the strong beam is ignored. Therefore the beam-beam effect on the particle trajectories of the weak beam may be treated in the single particle picture, i.e., as a nonlinear kick due to the electromagnetic forces experienced while passing through a (longitudinally) short and time-independent external charge distribution.

We ignore coupling to the longitudinal motion, and we assume that the strong beam is represented by an axially symmetric charge distribution around the common closed orbit of the two beams in the transverse coordinate plane, so that it suffices to study a single phase plane. We start by stating the model in the so-called canonical accelerator coordinates $v \equiv (v_1, v_2)^{\mathrm{T}}$, where v_1 has the dimension of a length and $v_2 := p_{v_1}/p_0$ is dimensionless (p_0 is the constant longitudinal momentum of the particle on the closed orbit, usually much larger than p_{v_1}, the canonical conjugate of v_1). Normally the lattice is chosen so that the unperturbed beam envelope at the interaction point has a local minimum, and thus the linear lattice is represented by $M := \begin{pmatrix} \cos(2\pi Q_0) & \beta\sin(2\pi Q_0) \\ -\sin(2\pi Q_0)/\beta & \cos(2\pi Q_0) \end{pmatrix}$, where $Q_0 \in \mathbf{R}$ and $\beta > 0$ are the unperturbed tune and the unperturbed beta-function of the weak beam at the interaction point, respectively (the beam envelope has width of order $\sqrt{\beta}$). The beam-beam kick is given by $v_2 \mapsto v_2 - \eta K(v_1)$, where $\eta := 8\pi\frac{\sigma_1^{*2}}{\beta}\xi$, and where $K(v_1) := \frac{1}{v_1}\left(1 - \exp\left(-\frac{v_1^2}{2\sigma_1^{*2}}\right)\right)$. Here σ_1^* is the spatial standard deviation of the Gaussian representing the strong beam, and ξ is the (typically small) linear beam-beam tune shift parameter. Our difference equation in the accelerator coordinates now reads $v_{n+1} = M\,v_n + \eta\,M\,(0, -K(v_{1,n}))^{\mathrm{T}}$.

Remark 3.1 In the special case of two matched, axially symmetric Gaussian beams, ξ is given by $\xi = \pm N^* r_p \beta/(4\pi\gamma\sigma_1^2)$, where N^* is the number of particles in the strong beam, r_p is the so-called classical particle radius of the species, σ_1 is the spatial RMS beam width of the two beams, and $\gamma > 1$ is the Lorentz factor of the weak beam.

We now rescale the variables according to $w \equiv (w_1, w_2)^{\mathrm{T}} := (v_1/\sigma_1,\, \beta v_2/\sigma_1)^{\mathrm{T}}$, where σ_1 is the standard deviation of v_1 for the weak beam when matched to its unperturbed lattice (i.e., when the phase space density depends only on $v^{\mathrm{T}}\,B^{-1}\,v$, where $B := \mathrm{diag}(\beta, 1/\beta)$ is the beam matrix at the interaction point; note that $\sigma_2 := \sigma_1/\beta$ is then the standard deviation of v_2 for the weak beam). In the rescaled variables the difference equation becomes

$$w_{n+1} = R\,w_n + \varepsilon\,R\begin{pmatrix} 0 \\ -H'(w_{1,n}) \end{pmatrix}, \tag{3.13}$$

where $R := e^{\mathcal{J}2\pi Q_0}$, $\varepsilon := 8\pi r^2\xi$, $r := \sigma_1^*/\sigma_1$, and $H'(w_1) := \dfrac{1}{w_1}\left(1 - \exp\left(-\dfrac{w_1^2}{2r^2}\right)\right)$. Thus Eq. (3.13) has the form of Eq. (3.1). We note that ε is dimensionless and small whenever ξ is small, that w_1 and w_2 are dimensionless and $O(1)$ for a typical particle trajectory of the weak beam, and that in a collider the two beams are typically matched to each other so that $r \approx 1$. By using the substitution $s^2/(2r^2) = w_1^2/(2r^2 + s')$ one can show that

$$H(w_1) \;:=\; \int_0^{w_1}\left(1 - \exp\left(-\frac{s^2}{2r^2}\right)\right)\frac{ds}{s} \;=\; \frac{1}{2}\int_0^{\infty}\left(1 - \exp\left(-\frac{w_1^2}{2r^2 + s'}\right)\right)\frac{ds'}{2r^2 + s'}\,, \tag{3.14}$$

where we have taken $H(0) = 0$.

Before proceeding we check the linearized behavior about the equilibrium $w = 0$. The linearization of Eq. (3.13) is $w_{n+1} = G\,w_n$, $G := R\begin{pmatrix} 1 & 0 \\ -4\pi\xi & 1 \end{pmatrix}$, where we have used the fact that $H''(0) = (2r^2)^{-1}$. The system is linearly stable if and only if $|\mathrm{tr}\,G| < 2$, i.e., provided the linearly perturbed tune Q, defined by $\cos(2\pi Q) := \frac{1}{2}\mathrm{tr}\,G = \cos(2\pi Q_0) - 2\pi\xi\,\sin(2\pi Q_0)$, is real and satisfies $|\cos(2\pi Q)| < 1$. It follows that $Q = Q_0 + \xi + O(\xi^2)$, thus justifying the name “linear beam-beam tune shift parameter” for ξ. For $Q_0 \in \{0, 1/2, 1\}$ we see that $|\mathrm{tr}\,G| = 2$, consistent with the discussion in the paragraph immediately following Eq. (3.1). For $Q_0 \in \{1/4, 3/4\}$, $|\mathrm{tr}\,G| = 2\pi|\xi|$ and thus we have linear stability, which is consistent with the results of Subsection 3.2.2.

3.2.1 The weak-strong beam-beam effect in the far-from-low-order-resonance case

For $Q_0 \in \mathcal{D}(\varepsilon^\lambda\phi, R_\varepsilon)$ the motion is given by Eq. (3.7), where $\nu \equiv Q_0$, and where ω is determined as follows. We use Eq. (3.14) to obtain $H_0(J) := \int_0^1 H(\sqrt{2J}\sin(2\pi t))\,dt = \frac{1}{2}\int_0^{J/(2r^2)}(1 - e^{-w}I_0(w))\,\frac{dw}{w}$, where I_0 is the zero-th order modified Bessel function and where we have used the expansion $\exp(x\cos(y)) = I_0(x) + 2\sum_{k=1}^{\infty} I_k(x)\cos(ky)$. Omega is given by

$$2\pi\omega(J) \;:=\; H_0'(J) \;=\; \frac{1}{2J}\left(1 - \exp\left(-\frac{J}{2r^2}\right)I_0\left(\frac{J}{2r^2}\right)\right) \;=\; \frac{1}{4\pi J}\int_0^{2\pi}\left(1 - \exp\left(-\frac{J\sin^2\vartheta}{r^2}\right)\right)d\vartheta\,. \tag{3.15}$$

The amplitude-dependent tune shift $\varepsilon\omega(J_0)$ is identical to that derived in [ES] and justifies the use of the delta function there. Notice also that $\varepsilon\omega(0) = \xi$, in agreement with the linearization above.

3.2.2 The weak-strong beam-beam effect in the near-to-low-order-resonance case

In Subsection 3.1.2 we found the Hamiltonian for the autonomous system (3.11) to be $\mathcal{K}(\check{z}) = 2\pi a J(\check{z}) + \widehat{\mathcal{H}}(\check{z}, 0)$, where $\widehat{\mathcal{H}}(\check{z}, 0) = (1/p)\sum_{n=0}^{p-1} H(\check{z}_1 \cos[2\pi nq/p] + \check{z}_2 \sin[2\pi nq/p])$ and $Q_0 = q/p + a\varepsilon$. Since $H(x)$ approaches zero for large x, $\mathcal{K}(\check{z})$ approaches $2\pi a J(\check{z})$, and for $a \neq 0$ the integral curves become circles at large distances from the origin. The motion on these circles is clockwise for positive a and counterclockwise for negative a, thus a bifurcation in the phase plane portrait occurs at $a = 0$. In the case where $q/p \in \{0, 1/2, 1\}$ it is easy to see that $\widehat{\mathcal{H}}(\check{z}, 0) = H(\check{z}_1)$, and for $q/p \in \{1/4, 3/4\}$ one also easily finds $\widehat{\mathcal{H}}(\check{z}, 0) = 1/2\,[H(\check{z}_1) + H(\check{z}_2)]$ since H is an even function. For $q/p \in \{1/3, 2/3\}$ we find $\widehat{\mathcal{H}}(\check{z}, 0) = 1/3\,[H(\check{z}_1) + H(-\check{z}_1/2 + \sqrt{3}\check{z}_2/2) + H(-\check{z}_1/2 - \sqrt{3}\check{z}_2/2)]$. We briefly discuss the phase plane portraits for $\mathcal{K}$ in these cases (see [DEV] for figures).

In the first case ($q/p \in \{0, 1\}$) and for $a = 0$ we have $d\check{z}_1/dt = 0$ and $d\check{z}_2/dt = H'(\check{z}_{1,0})$. Thus the motion is identical to the exact case, as discussed just before Eq. (3.2), since Eqs. (3.1) and (3.3) and the associated averaged problem are identical. For a small but positive, the origin is a (nonlinearly) stable center and the phase portrait is a one-parameter family of ovals which are long and thin in the $\check{z}_2$ and $\check{z}_1$ directions respectively. As a increases to modest values the ovals become circular, consistent with the expectation of "stability far from low-order resonance." As a decreases from zero, the origin becomes a saddle, and two centers emerge from infinity at $(\pm c, 0)$, where $c \sim 1/\sqrt{2\pi|a|}$ for $|a|$ small. As a decreases further, the centers coalesce with the saddle at $4\pi a r^2 = -1$, and for $4\pi a r^2 < -1$ the only critical point is a center at the origin, again consistent with our expectation of stability.

The motion for $q/p = 1/2$ in the period two Poincaré map is identical with the motion for $q/p = 1$; the intermediate values may be obtained by rotating the phase plane portrait by a half turn.

For $q/p \in \{1/4, 3/4\}$ the phase plane portrait has a four-fold symmetry, being invariant under reflections about the two axes and about the lines $\check{z}_2 = \pm\check{z}_1$. The origin is a critical point and its linearized vector field has eigenvalues $\pm 2\pi i(a - a_c)$, where $a_c = -1/(8\pi r^2)$. Thus the origin is a (nonlinearly) stable center for $a \neq a_c$, and it is easily checked that the origin is also a stable center for $a = a_c$ and that the rotation is clockwise for $a > a_c$ and counterclockwise for $a \leq a_c$. For $a \geq 0$ there are no other equilibria and the phase plane portrait is a one-parameter family of concentric ovals. For a small the (closed) integral curves look like four-pointed stars, with smoothed points on the axes, and as a increases the curves become circles. For $a_c < a < 0$ there are eight nonzero critical points. The four critical points $(\pm c, \pm c)$ are centers and the four at $(0, \pm c)$ and $(\pm c, 0)$ are saddle points, where c is the unique positive root of $4\pi a c + H'(c) = 0$. The critical points form an island structure in a neighborhood of radius c of the origin in the phase plane. This island structure emerges from infinity as a decreases through zero and coalesces into the origin as a decreases to a_c. For $a \leq a_c$, the origin is again the only equilibrium, and it is a stable center with counterclockwise rotation. The portrait is again a one-parameter family of ovals approaching circles as a decreases from a_c.

Because H is an even function, $\widehat{\mathcal{H}}$ is the same for all $q/p \in \{1/6, 1/3, 2/3, 5/6\}$. Thus the phase plane portraits are the same for resonances of order three and six, and these portraits have a six-fold symmetry, being invariant under reflections about the axes $\check{z}_1 = 0$, $\check{z}_2 = 0$ and the lines $\check{z}_2 = \pm\check{z}_1/\sqrt{3}$ and $\check{z}_2 = \pm\sqrt{3}\,\check{z}_1$. Qualitatively, the behavior as a function of a is similar to that in the case of resonance of order four (e.g., the island structure is similar, but there are now six rather than four islands). The critical value a_c at which the islands coalesce in the origin turns out to be the same as in the case $p = 4$.

3.2.3 Summary of the weak-strong beam-beam effect

Our basic equation is Eq. (3.13) with R and H' defined there. Remark 2.5 and the summary in Subsection 3.1.3 apply. Here we emphasize that the motion depends only on ξ (or equivalently ε), and on the fractional part of Q_0, and that we have a fairly complete description for $Q_0 \in [0, 1]$ over $O(1/\xi)$ time intervals. Away from low-order resonances, the motion is given by Eq. (3.7), with $\nu = Q_0$ and with ω defined in Eq. (3.15). Thus the motion takes place approximately on circles with an amplitude-dependent tune. Near low-order resonances the behavior is given by Eq. (3.12), and $\check{z}(t)$ evolves according to the time-independent Hamiltonian $\mathcal{K}(\check{z}) := 2\pi a J(\check{z}) + (1/p)\,\sum_{n=0}^{p-1} H(\check{z}_1 \cos(2\pi nq/p) + \check{z}_2 \sin(2\pi nq/p))$. As described above in Subsection 3.2.2, this Hamiltonian has a rich variety of behaviors depending on the order p of the resonance, and on the displacement $a\varepsilon$ from the resonance. In particular the behavior varies considerably for $a > 0$, $a = 0$ and $a < 0$. Finally, we again emphasize that while our description is fairly complete, there are gaps between the regions of validity of the nonresonant normal form which does not depend on Q_0, and the resonant normal form which does depend on Q_0 (cf. Remark 2.5).

3.3 The Hénon Map

We now apply Theorem 3 to the Hénon map (in beam dynamics this map is a standard model for the effect of a localized sextupole magnet in an otherwise linear lattice). The standard form of the Hénon map is Eq. (3.1) with $H(w_1) = w_1^3/3$. This gives Eq. (3.4) with $\mathcal{H}(x,\theta) = (x_1 \cos 2\pi\theta + x_2 \sin 2\pi\theta)^3/3$, which clearly has zero average. It follows that $f(x,\theta) = \mathcal{J}\,\nabla_x \mathcal{H}(x,\theta)$ in Eq. (1.4) has zero average, so that hypothesis (jw) of Theorem 3 is satisfied. Thus, by Theorem 3 and Proposition B, for appropriate $\varepsilon, T > 0$, $\nu \in D(\phi, R_\varepsilon)$, and for any $0 \le \alpha \le 1$, we have $|x_n - x_0| = O(\varepsilon^\alpha)$ on the discrete time interval $0 \le n \le T/\varepsilon^{2-\alpha}$.

Remark 3.2 The above discussion simply applies Theorem 3 as is (and thus also covers the case of more general H), but when $\mathcal{H}$ has a finite Fourier series (e.g. when H is a polynomial, as above) the proof of Theorem 3 may be simplified, both in terms of the smoothness requirement (see Remark 4.4) and in terms of the estimates in Lemma 2. In particular, for the Hénon map above, $g_k = 0$ except for $|k| \in \{1, 3\}$, so taking $R_\varepsilon = 3$, we see that the series defining C_1 and C_2 in Lemma 2 have only four terms each, while the tail-series of Lemma 2 vanishes.

4. Proofs and Additional Mathematical Results

As the title indicates, this is the most mathematical section of the paper. Subsection 4.1 treats periodic maps; this is quite straightforward, and may be read as a kind of introduction to the deeper results of the next subsection. Subsection 4.2 concerns the considerably more complex case of maps far from low-order resonance, and requires a (short) discussion of small divisors and truncated Diophantine conditions. The use of such conditions is not new (for example, related conditions are used to obtain general multiphase averaging results in [ABG]), but we believe our use of them in the present context is the most innovative aspect of this paper.

4.1 Periodic Systems

In this subsection we give a self-contained presentation of the remarkably simple technology required to prove the averaging principle for maps with periodic perturbations. This consists of the Besjes inequality for periodic functions (below), followed by its application to the proof of Theorem 1.

4.1.1 The Besjes inequality for periodic functions

Let $U \subset \mathbf{R}^d$ be open, and $S = U \times \mathbf{N}$. The Besjes inequality relies in an essential way upon the following assumption concerning the function $g : S \to \mathbf{R}$, periodic with period p in its second argument:

(iv) For each $x \in U$, $\displaystyle\sum_{n=0}^{p-1} g(x,n) = 0$

When g has period p in n and satisfies (iv), we say it has *zero mean in n*. We now state the Besjes inequality for periodic maps as

Lemma 1. *Let $U \subset \mathbf{R}^d$ be open, $S = U \times \mathbf{N}$, and suppose $g : S \to \mathbf{R}$ satisfies assumptions (i), (ii) (from §2.1) and (iv) above and is globally x-Lipschitz with Lipschitz constant $L \ge 0$. If $\{x_n\}_{n=0}^{\infty} \subset U$ is a sequence for which the successive differences $x_{n+1} - x_n$ are bounded by M (i.e., $\sup_n |x_{n+1} - x_n| \le M$), then for all $N \in \mathbf{N}$,*

$$\Big|\sum_{n=0}^{N-1} g(x_n, n)\Big| \;\le\; \frac{1}{2}\, NpLM \;+\; p\,\|g\|_S .$$

Proof. Using the notation $[a]$ to designate the greatest integer in a, we first set $l = [(N-1)/p]$ (so that l is the number of periods of g contained in the segment $\{0, 1, 2, \dots, N-1\}$). Then using the fact that g is periodic and of zero mean, we write

$$\sum_{n=0}^{N-1} g(x_n, n) \;= \sum_{k=0}^{l-1} \sum_{n=0}^{p-1} \Big(g(x_{n+kp}, n) - g(x_{kp}, n) \Big) \;+\; \sum_{n=lp}^{N-1} g(x_n, n)\,.$$

Now since g is Lipschitz in its first argument, and since $|x_{n+kp} - x_{kp}| \le Mn$, we have

$$\Big|\sum_{n=0}^{N-1} g(x_n, n)\Big| \;\le\; \sum_{k=0}^{l-1}\sum_{n=0}^{p-1} LMn \;+\; \sum_{n=lp}^{N-1} |g(x_n,n)|$$

$$\le\; lLM\,\frac{p(p-1)}{2} \;+\; p\,\|g\|_S \;\le\; \frac{1}{2}NpLM \;+\; p\,\|g\|_S\,. \quad //$$

Remark 4.1 The original version of this lemma (Lemma 1 of [Bes]) was formulated for use in the proof of averaging principles for ODEs on $O(1/\varepsilon)$ timescales, and we use its analog in a similar way below for maps. The original lemma bounds the time by a constant that is $O(1/\varepsilon)$ and gives a final bound that is $O(\varepsilon)$, independent of time. We have found, however, that retaining the (here discrete) time-dependence makes the result more versatile (cf. the proof of Theorem 3 below).

Remark 4.2 Lemma 1 (and many of its generalizations) may also be proved using "summation by parts," as in the proof of Lemma 2 below.

We now illustrate the use of Lemma 1 by using it to prove Theorem 1.

4.1.2 Proof of Theorem 1

Assume the hypotheses of Theorem 1 (cf. §2.1). It is clear from assumption (iii) that the solutions x_N and y_N exist uniquely for all $N \in \mathbf{N}$. To see that the approximation relation holds, we write

$$|x_N - y_N| \;=\; \varepsilon\Big|\sum_{n=0}^{N-1}\Big(f(x_n,n) - \widehat{f}(y_n)\Big)\Big| \;=\; \varepsilon\Big|\sum_{n=0}^{N-1}\Big(f(x_n,n) - f(y_n,n) + f(y_n,n) - \widehat{f}(y_n)\Big)\Big|$$

$$\le\; \varepsilon L \sum_{n=0}^{N-1} |x_n - y_n| \;+\; \varepsilon\Big|\sum_{n=0}^{N-1}\widetilde{f}(y_n,n)\Big|$$

where L is the x-Lipschitz constant of f, and where $\widetilde{f}(y,n) := f(y,n) - \widehat{f}(y)$ (the "oscillating part of f") satisfies the hypotheses of Lemma 1 with $U = \mathbf{R}^d$ (in particular, $\widetilde{f}$ has zero mean and y-Lipschitz constant $2L$). Using the fact (from Eq. (1.2) and assumption (i)) that $|y_{n+1} - y_n| \le M := \varepsilon\|\widehat{f}\|_{\mathbf{R}^d}$, we have $|x_N - y_N| \le \varepsilon\, L\sum_{n=0}^{N-1}|x_n - y_n| + \varepsilon\,\frac{1}{2}\,N\,p\,2L\,\varepsilon\|\widehat{f}\|_{\mathbf{R}^d} + \varepsilon\, p\,\|\widetilde{f}\|_S$. Thus $|x_N - y_N| \;\le\; \varepsilon L\sum_{n=0}^{N-1}|x_n - y_n| + \varepsilon\, p\,\big(LT\|\widehat{f}\|_{\mathbf{R}^d} + \|\widetilde{f}\|_S\big)$ for $0 < N \le T/\varepsilon$. Applying Gronwall's inequality for sequences (Lemma 3 in the Appendix) and setting $C = \big(LT\|\widehat{f}\|_{\mathbf{R}^d} + \|\widetilde{f}\|_S\big)e^{LT}$ gives $|x_N - y_N| \;\le\; Cp\,\varepsilon$ for $0 < N \le T/\varepsilon$, as claimed. The second part of Theorem 1 (namely $|x_n - y(n)| \le (Cp + C')\varepsilon$ for $0 \le n \le T/\varepsilon$) follows from Lemma 4 (Appendix) and the triangle inequality. //

Remark 4.3 The preceding is no doubt one of the simplest possible proofs of an averaging principle for maps. Part of the simplicity derives from the use of Lemma 1, and part derives from the assumption of compact support (iii), which permits us to dispense with questions of the existence intervals for solutions. Thus, although assumption (iii) is often invalid in practice, by using it we are able to show that the basic estimates of the averaging method do not require restrictions on the size of ε; such restrictions are instead introduced by considering solutions' existence intervals, or by methods of proof which rely on near-identity transformations (which may in turn require restrictions on ε for their inversion). Of course our results may be extended to cases with finite existence intervals (see Proposition B, §2.4), and may also be combined with more traditional transformation methods to obtain efficient results at higher order [DESV].

4.2 Systems Far From Low-Order Resonance

In this subsection we generalize the Besjes inequality to functions far from low-order resonance in their second argument. We then use this inequality to prove Theorems 2 and 3. First, however, we present the following brief discussion.

4.2.1 Resonant zones, Diophantine conditions, and the ultraviolet cutoff

Before stating and proving our next analog of Besjes' inequality, we discuss aspects of resonance, small divisors and Diophantine conditions that will be needed in the sequel. A more comprehensive introduction may be found in [Yo].

Zone Functions and Diophantine Conditions

In dynamical systems, Diophantine conditions arise naturally as a means of "controlling small divisors" and "avoiding resonances." Typically, in one dimension, divisors of the form $e^{2\pi i k\nu} - 1$ (with $0 \neq k \in \mathbf{Z}$ and $0 \neq \nu \in \mathbf{R}$) occur as the denominators of terms in a series indexed over k, together with numerators which decrease to zero with increasing $|k|$. Clearly divisors cannot vanish, so rational (or "resonant") values of ν must be avoided. And although irrational ν do not cause divisors to vanish, when "nearly resonant," they may generate such small divisors as to cause divergence of the series in which they occur.

By using a suitably decreasing *zone function* $\phi : \mathbf{R}_+ \to \mathbf{R}_+$ (the inverse of which is called an "approximation function" in [Rü]), we define the "highly nonresonant" values of ν as those belonging to the corresponding *Diophantine set*

$$\mathcal{D}(\phi) = \{\nu \in \mathbf{R} \,\big|\, |e^{2\pi i k\nu} - 1| \geq \phi(|k|),\ k \in \mathbf{Z}\backslash\{0\}\}, \tag{4.1}$$

which is a Cantor set. The Diophantine set $\mathcal{D}(\phi)$ may be thought of as $\mathbf{R}$ with countably many *zones* removed, where the zone $\mathcal{Z}_k = \{\nu \in \mathbf{R} \,\big|\, |e^{2\pi i k\nu} - 1| < \phi(|k|)\}$ corresponding to a particular $k \neq 0$ is the countable union of open intervals centered on rational numbers of the form q/k ($q \in \mathbf{Z}$). To better see the structure of $\mathcal{D}(\phi)$, consider its intersection with the interval $[0,1]$. For each fixed $k > 0$ we remove k intervals of length 2δ from $[0,1]$, where $|e^{i2\pi k(\delta + l/k)} - 1| = |e^{i2\pi k\delta} - 1| < \phi(k)$. For small $\phi(k)$, this gives $\delta \approx \phi(k)/(2\pi k)$, and thus the total length of $\mathcal{Z}_k \cap [0,1]$ is $2\delta k \approx \phi(k)/\pi$. It follows that the total length of the union $\cup_k \mathcal{Z}_k \cap [0,1]$ of the overlaps of all zones $\mathcal{Z}_k$ with $[0,1]$ is (approximately) bounded by $\sum_k \mathrm{length}(\mathcal{Z}_k \cap [0,1]) \approx (1/\pi)\int_1^\infty \phi(k)\,dk$. Thus a typical zone function of the form $\phi(r) = \gamma r^{-(\tau+1)}$ with $\tau > 0$ removes zones of total length no more than $\gamma/(\pi\tau)$ from $[0,1]$. When this total length is less than one, the Diophantine set $\mathcal{D}(\phi)$ has positive measure (and is therefore nonempty).

More generally, if the zone function ϕ decreases too slowly, then the union of the excluded zones may be so large that its complement, $\mathcal{D}(\phi)$, is empty. Conversely, if ϕ decreases too rapidly, then $\mathcal{D}(\phi)$ may be too large, and may contain values of ν so close to resonance as to cause divergence of the series in which small divisors appear.

The following terminology is useful for describing zone functions that permit convergence of the series arising in the proof of Lemma 2 below. If $U \subset \mathbf{R}^d$ is open, and $f : U \times \mathbf{R} \to \mathbf{R}^d$ has period 1 in its second argument and Fourier series $f(x,\theta) \sim \sum_{k\in\mathbf{Z}} f_k(x)\, e^{2\pi i k\theta}$ (where the kth Fourier coefficient is $f_k(x) = \int_0^1 f(x,\theta)\, e^{-2\pi i k\theta}\, d\theta$, requiring only that f is integrable in θ), then given a zone function ϕ such that $\mathcal{D}(\phi) \neq \emptyset$, we say that ϕ is *adapted to f on U* provided

$$\sum_{0\neq k\in\mathbf{Z}} \frac{\|f_k\|_U}{\phi(|k|)} < \infty \qquad \text{and} \qquad \sum_{0\neq k\in\mathbf{Z}} \frac{\|Df_k\|_U}{\phi(|k|)} < \infty\,, \tag{4.2}$$

where Df_k denotes the derivative of the function $f_k : U \to \mathbf{R}^d$. Smoothness conditions on f assuring the existence of zone functions adapted to f are not severe, as we now show.

Smoothness Conditions Ensuring the Existence of Adapted Zone Functions

Several questions naturally arise concerning the relationship between the smoothness of f and the existence of zone functions adapted to f as in Eq. (4.2). Formulating the sharpest possible conditions in this direction is somewhat delicate, but the following brief discussion should serve as a good starting point.

We first recall that for $\tau > 0$, the zone function $\phi(r) = \gamma r^{-(\tau+1)}$ generates a nonempty Diophantine set $\mathcal{D}(\phi)$ provided $\gamma > 0$ is sufficiently small (see the preceding discussion, or the more extensive discussion in §1.2 of [BHS]). We assume that $f : U \times \mathbf{R} \to \mathbf{R}^d$ is of class $C^{p+1}(U \times \mathbf{R})$ and of compact support in the first argument, uniformly with respect to the second (cf. assumption (jjj) in §2.2). Integrating the kth Fourier coefficient $f_k(x) = \int_0^1 f(x,\theta)\, e^{-2\pi i k\theta}\, d\theta$ by parts p times with respect to θ gives $f_k(x) = (2\pi i k)^{-p} \int_0^1 \left[\partial^p f/\partial\theta^p\right](x,\theta) e^{-2\pi i k\theta}\, d\theta$. Then taking the supremum over $x \in U$ of both sides of this expression gives $\|f_k\|_U \le C(f,p)|k|^{-p}$, where $C(f,p) = \frac{1}{(2\pi)^p} \sup_{x\in U} \int_0^1 \left|\frac{\partial^p f}{\partial\theta^p}(x,\theta)\right| d\theta$. The same estimate holds for $\|Df_k\|_U$ with $C(f,p)$ replaced by $C'(f,p) = \frac{1}{(2\pi)^p} \sup_{x\in U} \int_0^1 \left|\frac{\partial^{p+1} f}{\partial x\partial\theta^p}(x,\theta)\right| d\theta$.

Using these estimates, we immediately deduce that both of the series in Eq. (4.2) are convergent provided that $p > \tau + 2$. Conversely, we see that whenever $p \ge 3$, there exists a zone function $\phi(r) = \gamma r^{-(\tau+1)}$ with $0 < \tau < p - 2$ which generates nonempty Diophantine sets $\mathcal{D}(\phi)$ (for γ sufficiently small) and which is adapted to f in the sense of Eq. (4.2). This justifies our assumption (j) in Theorems 2 and 3.

Remark 4.4 A more refined (and lengthy) argument shows that the existence of ϕ adapted to f does not require quite as much smoothness as we demand above; we start our discussion under the assumption $f \in C^{p+1}(U \times \mathbf{R})$ primarily for simplicity. Of course, when f has a (sufficiently short) finite Fourier series, the decay rate of its terms is not an issue, and the smoothness requirement may be reduced to C^1.

Remark 4.5 Although our results for system (1.4) as presented in this paper do not apply to the case of analytic perturbations εf (since analytic f with compact support vanishes identically), it would not be especially difficult to extend our theory to this case. For analytic $f : U \times \mathbf{T}^1 \to \mathbf{R}$ with Fourier coefficients f_k decreasing exponentially as, say, $\|f_k\|_U \le \Gamma e^{-\beta|k|}$, it would be appropriate to use exponentially decreasing zone functions, for which the preceding discussion is easily modified. In fact, given any $\rho > 0$, the zone function $\phi(r) = \gamma e^{-\rho r}$ generates nonempty Diophantine sets $\mathcal{D}(\phi)$ for small enough $\gamma > 0$. The decay rate β of the f_k must of course exceed ρ, which can be arranged provided f is analytic in its second argument with *analyticity parameter* $\alpha > \rho$ (this is an instance of the Paley-Wiener Lemma; cf. [PW] or [BHS]). Roughly speaking, the analyticity parameter α is a measure of the minimum distance by which f may be extended as an analytic function of the complex torus (see also §4.3.3 of [DEG] for an elementary discussion in the two-dimensional case).

It is interesting to note that Diophantine conditions corresponding to exponentially decaying zone functions ϕ may be strictly weaker than the weakest small-divisor conditions ordinarily used in dynamical systems, the so-called *Bruno conditions* (also spelled Brjuno or Bryuno; here "strictly weaker" means that the set $\mathcal{D}(\phi)$ properly contains the set of ν subject to Bruno conditions). This is however not surprising, since Bruno conditions apply to situations (such as conjugacies of circle diffeomorphisms, or KAM theory) in which countably many series with small divisors must simultaneously converge. By contrast, in Lemma 2 we require the convergence of only two series (in the language of [BHS], ours is a "one-bite" small-divisor problem).

The Ultraviolet Cutoff and Truncated Diophantine Conditions

Finally, we introduce the notion of ultraviolet cutoff, which is important in physical applications of Diophantine conditions. To understand why, note that typically in applications, the ν that are required to be Diophantine are physical parameters. But checking whether a given ν belongs to a Cantor set of the form $\mathcal{D}(\phi)$ is a practical impossibility, since each point of $\mathcal{D}(\phi)$ has points arbitrarily close to it that are not in $\mathcal{D}(\phi)$. In other words, deciding if ν belongs to $\mathcal{D}(\phi)$ requires ν to be specified with infinite precision. Practically of course, it is only possible to specify physical parameters with finite precision. We surmount this difficulty by introducing *truncated Diophantine conditions* of the form

$$\mathcal{D}(\phi, R) = \{\nu \in \mathbf{R} \,\big|\, |e^{2\pi i k\nu} - 1| \ge \phi(|k|),\ k \in \mathbf{Z} \text{ with } 0 < |k| \le R\}. \tag{4.3}$$

When $\nu \in \mathcal{D}(\phi, R)$, we say ν is *Diophantine to order R with respect to ϕ*, and we call R the *truncation order* or (*ultraviolet*) *cutoff*. Note that $\mathcal{D}(\phi, R)$ is an approximating superset of $\mathcal{D}(\phi)$ with nonempty interior which converges to $\mathcal{D}(\phi)$ as $R \to \infty$. To decide whether ν belongs to $\mathcal{D}(\phi, R)$, one checks only finitely many inequalities.

As a rough general rule, results in dynamical systems which are established for Diophantine sets $\mathcal{D}(\phi)$ may also be established (usually in slightly weaker form) for the corresponding larger, nicer sets $\mathcal{D}(\phi, R)$. The standard technique for doing so involves removing the "R-tail" of a series before applying Diophantine conditions, then checking that the tail is small. This technique was called the "ultraviolet cutoff" by Arnold in his proof of the KAM theorem [Ar], and is illustrated in the proof of Lemma 2 below.

4.2.2 Besjes' inequality generalized to functions far from low-order resonance

Lemma 2. *Let $S = \mathbf{R}^d \times \mathbf{R}$, and suppose $g : S \to \mathbf{R}^d$ satisfies assumptions (j), (jj) from Subsection 2.2, along with assumption (jw) from Subsection 2.3. Let the zone function ϕ be adapted to g on $\mathbf{R}^d$ in the sense of Eq. (4.2), and define the positive constants $C_1 = C_1(g, \phi)$ and $C_2 = C_2(g, \phi)$ by $C_1 = 2\sum_{0\neq k} \|g_k\|_{\mathbf{R}^d}/\phi(|k|)$ and $C_2 = \sum_{0\neq k} \|Dg_k\|_{\mathbf{R}^d}/\phi(|k|)$. Let $\nu \in \mathcal{D}(\phi, R)$. If $\{x_n\}_{n=0}^{\infty} \subset \mathbf{R}^d$ is a sequence for which the successive differences $x_{n+1} - x_n$ are bounded by M (i.e., $\sup_n |x_{n+1} - x_n| \leq M$), then*

$$\left|\sum_{n=0}^{N-1} g(x_n, n\nu)\right| \;\leq\; C_1 \,+\, N\Big(C_2 M + \sum_{|k|>R} \|g_k\|_{\mathbf{R}^d}\Big)\,, \qquad \text{where} \qquad \sum_{|k|>R} \|g_k\|_{\mathbf{R}^d} \to 0 \quad \text{as} \quad R \to \infty.$$

Proof. Since $C_1 < \infty$, we write g as its uniformly convergent Fourier series $g(x,\theta) = \sum_{0\neq k \in \mathbf{Z}} g_k(x) e^{2\pi i k\theta}$, so that

$$\left|\sum_{n=0}^{N-1} g(x_n, n\nu)\right| \;\leq\; \left|\sum_{n=0}^{N-1} \sum_{0<|k|\leq R} g_k(x_n) e^{2\pi i k n\nu}\right| \;+\; \left|\sum_{n=0}^{N-1} \sum_{|k|>R} g_k(x_n)\, e^{2\pi i k n\nu}\right|. \tag{4.4}$$

We shall treat separately each of the double sums on the right-hand side of inequality (4.4). For the first double sum, we reverse the order of summation and use the "summation by parts" formula $\sum_{n=0}^{N-1} a_n(b_{n+1} - b_n) = (a_N b_N - a_0 b_0) - \sum_{n=0}^{N-1}(a_{n+1} - a_n) b_{n+1}$ with $a_n = g_k(x_n)$ and $b_n = e^{2\pi i k n\nu}/(e^{2\pi i k\nu} - 1)$ so that $a_n(b_{n+1} - b_n) = g_k(x_n) e^{2\pi i k n\nu}$. It then follows that

$$\begin{aligned}
\left|\sum_{0<|k|\leq R} \sum_{n=0}^{N-1} g_k(x_n) e^{2\pi i k n\nu}\right| \;&\leq\; \sum_{0<|k|\leq R} \left| \frac{g_k(x_N) e^{2\pi i N k\nu} - g_k(x_0)}{e^{2\pi i k\nu} - 1} \;-\; \sum_{n=0}^{N-1} \big(g_k(x_{n+1}) - g_k(x_n)\big) \frac{e^{2\pi i (n+1) k\nu}}{e^{2\pi i k\nu} - 1} \right| \\
&\leq\; \sum_{0<|k|\leq R} \left(\frac{2\|g_k\|_{\mathbf{R}^d}}{|e^{2\pi i k\nu} - 1|} \;+\; \frac{\|Dg_k\|_{\mathbf{R}^d}}{|e^{2\pi i k\nu} - 1|} \sum_{n=0}^{N-1} |x_{n+1} - x_n| \right) \;\leq\; \sum_{0<|k|\leq R} \frac{2\|g_k\|_{\mathbf{R}^d} + NM\|Dg_k\|_{\mathbf{R}^d}}{|e^{2\pi i k\nu} - 1|} \\
&\leq\; \sum_{0<|k|\leq R} \frac{2\|g_k\|_{\mathbf{R}^d} + NM\|Dg_k\|_{\mathbf{R}^d}}{\phi(|k|)} \;\leq\; \sum_{0\neq k} \frac{2\|g_k\|_{\mathbf{R}^d} + NM\|Dg_k\|_{\mathbf{R}^d}}{\phi(|k|)} \;=\; C_1 \,+\, NMC_2\,.
\end{aligned} \tag{4.5}$$

We next treat the second double sum (the R-tail) on the right-hand side of inequality (4.4) using the simple estimate

$$\left|\sum_{n=0}^{N-1} \sum_{|k|>R} g_k(x_n)\, e^{2\pi i k n\nu}\right| \;\leq\; N \sum_{|k|>R} \|g_k\|_{\mathbf{R}^d} \;\to\; 0 \quad \text{as} \quad R \;\to\; \infty\,. \tag{4.6}$$

Inserting estimates (4.5) and (4.6) into inequality (4.4) concludes the proof. //

Remark 4.6 A related analogous result for flows (but without the ultraviolet cutoff) appears as Lemma 13 of [Sá], and in Theorem 2 of [ES], and a more general Besjes-type inequality for so-called KBM vector fields also appears in [Sá] as Lemma 2. A still more closely related result for flows appears as Lemma 2 in our previous paper [DEG], where it was used in averaging methods applied to certain classes of charged particle motions in crystals.

Remark 4.7 In the case where g has a finite Fourier series, the above proof simplifies in obvious ways; but these simplifications become problematic as the Fourier series grows in length (note that the example in §3.3 has a Fourier series with only four terms).

4.2.3 Proof of Theorem 2

Assume the hypotheses of Theorem 2 (cf. §2.2; note that assumption (j) ensures the existence of zone functions adapted to f, as discussed before Remark 4.4). The proof is essentially the same as the proof of Theorem 1 with appropriate changes as needed in order to use Lemma 2. As in the previous proof, the solutions x_N and y_N clearly exist uniquely for all $N \in \mathbf{N}$. For the approximation relation, we write as before

$$|x_N - y_N| \leq\; \varepsilon L \sum_{n=0}^{N-1} |x_n - y_n| \;+\; \varepsilon \left|\sum_{n=0}^{N-1} \widetilde{f}(y_n, n\nu)\right|$$

where $\widetilde{f}(y,\theta) := f(y,\theta) - \overline{f}(y)$ is the oscillating part of f. The hypotheses clearly imply that $\|f\|_S < \infty$, and since ϕ is adapted to f on $\mathbf{R}^d$, the constants C_1 and C_2 from Lemma 2 are well defined. We may thus set $C = (C_1 + C_2 T\|\overline{f}\|_{\mathbf{R}^d} + 1)e^{LT}$. Finally, we fix the parameter $\zeta > 0$ and choose $R_\varepsilon > 0$ so large that $\sum_{|k|>R_\varepsilon} \|f_k\|_{\mathbf{R}^d} \le \zeta\varepsilon$, where $f_k(x)$ is the kth Fourier coefficient of f. It is now a simple matter to check that if $\nu \in \mathcal{D}(\phi, R_\varepsilon)$, then the hypotheses of Lemma 2 are satisfied with $M := \varepsilon\|\overline{f}\|_S$. We thus have

$$|x_N - y_N| \;\le\; \varepsilon\, L \sum_{n=0}^{N-1} |x_n - y_n| \;+\; \varepsilon C_1 \;+\; \varepsilon N\Big(C_2 M + \sum_{|k|>R} \|f_k\|_{\mathbf{R}^d}\Big)$$

$$\le\; \varepsilon\, L \sum_{n=0}^{N-1} |x_n - y_n| \;+\; \varepsilon C_1 \;+\; \varepsilon N\big(C_2 \varepsilon \|\overline{f}\|_{\mathbf{R}^d} + \zeta\varepsilon\big)\,,$$

and so for $0 < N \le T/\varepsilon$, we have $|x_N - y_N| \;\le\; \varepsilon L \sum_{n=0}^{N-1} |x_n - y_n| \;+\; \varepsilon\big(C_1 + C_2 T\|\overline{f}\|_{\mathbf{R}^d} + \zeta\big)$. Applying Gronwall's inequality for sequences (Lemma 3, Appendix) gives $|x_N - y_N| \;\le\; \varepsilon\big(C_1 + C_2 T\|\overline{f}\|_{\mathbf{R}^d} + \zeta\big)e^{\varepsilon LN} \;\le\; \varepsilon\big(C_1 + C_2 T\|\overline{f}\|_{\mathbf{R}^d} + \zeta\big)e^{LT} \;=\; C\varepsilon$ for $0 < N \le T/\varepsilon$, as claimed. The second part of Theorem 2 (namely $|x_n - y(n)| \le C'\varepsilon$ for $0 \le n \le T/\varepsilon$) again follows from Lemma 4 (Appendix) and the triangle inequality. //

Remark 4.8 It is important to note that for fixed positive ζ and ε, the ultraviolet cutoff R_ε need not be very large to ensure that $\sum_{|k|>R_\varepsilon} \|f_k\|_{\mathbf{R}^d} \le \zeta\,\varepsilon$, whence the number of inequalities to be checked in Eq. (4.3) (with $R = R_\varepsilon$) is also modest. In fact, straightforward estimation shows that when the Fourier coefficients of f decrease as $\|f_k\|_{\mathbf{R}^d} \le C|k|^{-(p+1)}$ (e.g. when f is of class C^{p+1}), it is enough to take $R_\varepsilon \ge 1 + \big(\frac{2C}{p\zeta\varepsilon}\big)^{1/p}$ (and when the coefficients decrease as $\|f_k\|_{\mathbf{R}^d} \le Ce^{-\rho|k|}$, it is enough to take $R_\varepsilon \ge 1 + \ln\big(\frac{2C}{\rho\zeta\varepsilon}\big)^{1/\rho}$).

Remark 4.9 If an $O(\varepsilon^2)$ term is added to Eq. (1.4) so that it reads $x_{n+1} \;=\; x_n + \varepsilon f(x_n, n\nu) + \varepsilon^2 g(x_n, n\nu)$, where $g : S \to \mathbf{R}^d$ satisfies the hypotheses of Theorem 2, then it is a simple matter to check that Theorem 2 continues to hold with the order constant C replaced by $C' = (C_1 + C_2 T\|\overline{f}\|_{\mathbf{R}^d} + \zeta + \|g\|_S)e^{LT}$. This form of Theorem 2 is often useful in applications.

4.2.4 Proof of Theorem 3

Assume the hypotheses of Theorem 3 (these include those of Theorem 2 together with the additional zero-mean assumption (jw); cf. §2.3). The hypotheses clearly imply that $\|f\|_S < \infty$, and since ϕ is adapted to f on $\mathbf{R}^d$, the constants C_1 and C_2 from the conclusion of Lemma 2 are well defined. We may thus choose the parameter $\zeta > 0$ and set $K_1 = C_1$ and $K_2 = C_2\|f\|_S + \zeta$. Finally, we choose $R_\varepsilon > 0$ so large that $\sum_{|k|>R_\varepsilon} \|f_k\|_{\mathbf{R}^d} \le \zeta\,\varepsilon$. It is now a simple matter to check that whenever $\nu \in \mathcal{D}(\phi, R_\varepsilon)$, the hypotheses of Lemma 2 are satisfied with $M := \varepsilon\|f\|_S$, from which we conclude that

$$|x_N - x_0| \;=\; \varepsilon\left|\sum_{n=0}^{N-1} f(x_n, n\nu)\right| \;\le\; \varepsilon C_1 + \varepsilon N\Big(C_2 M + \sum_{|k|>R_\varepsilon} \|f_k\|_{\mathbf{R}^d}\Big)$$

$$\le\; \varepsilon\, C_1 + \varepsilon N\big(C_2\varepsilon\|f\|_S + \zeta\varepsilon\big) \;\le\; K_1\varepsilon + K_2\varepsilon^2 N. \quad //$$

Remark 4.10 The proof of Theorem 3 is so short, and its hypotheses are so closely related to those of Lemma 2, that it is nearly a corollary of Lemma 2. The interesting features of Theorem 3 are that long-time invariance is shown without the traditional transformation of variables, while ν is required to be Diophantine only to low order R_ε.

4.3 Proofs of Propositions A, B, and C

For the statements of Propositions A, B, and C, see Subsection 2.4.

4.3.1 Proof of Proposition A

The zone functions enter the proofs of Theorems 2 and 3 only through Lemma 2. It is clear that if ϕ is replaced by $\varepsilon^\lambda\phi$ in Eq. (4.5), then the final estimate of Lemma 2 is changed to $\varepsilon^{-\lambda}(C_1 + NMC_2)$. The error bound in Theorem 2 then changes to $|x_N - y_N| \le \varepsilon^{1-\lambda}(C_1 + C_2 T\|\overline{f}\|_{\mathbf{R}^d})e^{LT} + \varepsilon\zeta e^{LT} = O(\varepsilon^{1-\lambda})$ for $0 \le N \le T/\varepsilon$, while the error bound in Theorem 3 changes to $|x_N - x_0| \le \varepsilon^{1-\lambda}C_1 + \varepsilon N\big(C_2\varepsilon^{1-\lambda}\|f\|_S + \zeta\varepsilon\big) \le K_1\varepsilon^{1-\lambda} + K_2\varepsilon^{2-\lambda}N$. //

4.3.2 Proof of Proposition B

Here we give the proof of Proposition B as it applies to Theorem 1 only; the proofs of its applicability to Theorems 2 and 3 are nearly the same.

Fix $\varepsilon > 0$, let $U \subset \mathbf{R}^d$ be open, take $S' = U \times \mathbf{N}$, and suppose $g : S' \to \mathbf{R}^d$, where g is not assumed to have compact support in U (in other words, g satisfies assumptions (i) and (ii) of §2.1, with S' in place of S, but does not satisfy assumption (iii)).

We now use g to define the systems ($1'$), ($2'$), and ($3'$), which are simply the previous systems (1.1), (1.2), and (1.3), respectively, in which the perturbation εf has been replaced by εg. We assume that the common initial condition $x_0 = y_0 = y(0)$ is fixed in U, and we choose the positive timescale parameter $T < \beta(x_0)$, where $[0, \beta(x_0))$ is the maximal forward interval of existence for the initial value problem

$$\frac{d\check{y}}{dt'} = \widehat{g}(\check{y}) \,, \qquad \check{y}(0) = x_0 \in U, \tag{$3''$}$$

which is simply the scaled, ε-independent version of system ($3'$) obtained by introducing the "slow time" $t' = \varepsilon t$. We then let $Z = \{z \in U \,|\, z = \check{y}(t'),\ 0 \le t' \le T\}$ denote the solution curve of system ($3''$) over $[0,T]$, and we choose $\delta > 0$ such that $\delta < \operatorname{dist}(Z, \partial U)$. Then the closure $\overline{D}(\delta)$ of the open "δ-tube" $D(\delta)$ around Z formed by the union of open balls of radius δ having centers in Z is contained in U; i.e., $D(\delta) := \bigcup_{t\in[0,T]} B_\delta\big(y(t)\big) \ \subset\ \overline{D}(\delta) \ \subset\ U$, where $B_\delta(y)$ denotes the open ball of radius δ centered on y in $\mathbf{R}^d$.

We next choose $r > 0$ so that the open ball $B_r(x_0)$ contains $\overline{D}(\delta)$, and we define the compactly supported function $f : \mathbf{R}^d \times \mathbf{N} \to \mathbf{R}^d$ which (a) coincides with g on $\overline{D}(\delta) \times \mathbf{N}$, (b) vanishes on $B_r(x_0)^{\mathrm{c}} \times \mathbf{N}$ (here $^{\mathrm{c}}$ denotes "complement"), and (c) interpolates g on $B_r(x_0) \cap \overline{D}(\delta)^{\mathrm{c}}$ in such a way that f is of the same smoothness class as g and such that $\|f\|_{\mathbf{R}^d\times\mathbf{N}} = \|g\|_{\overline{D}(\delta)\times\mathbf{N}}$. The existence of such f is guaranteed by the "smooth Tietze extension theorem" as given, for example, on p. 380 of [AMR].

Using this f, and the constant T (from the existence interval $0 \le t' \le T$ of the solution $\check{y}(t')$ of ($3''$), corresponding to the existence interval $0 \le t \le T/\varepsilon$ for the solution $y(t) = \check{y}(\varepsilon t)$ of ($3'$)), we apply Theorem 1 and Lemma 4 from the Appendix to conclude that, for appropriate $C_1, C_2 > 0$, we have:

$$|x_n - y_n| \le C_1\,\varepsilon \qquad \text{for} \qquad 0 \le n \le T/\varepsilon\,, \quad y_n \in D(\delta/2)\,, \quad \text{and} \quad x_n \in D(\delta)\,; \qquad \text{and}$$

$$|y_n - \check{y}(\varepsilon n)| \le C_2\,\varepsilon \qquad \text{for} \qquad 0 \le n \le T/\varepsilon \quad \text{and} \quad y_n \in D(\delta/2)\,.$$

Using these inequalities together with the triangle inequality, if we now impose a smallness condition on ε by requiring it to be strictly less than the threshold $\varepsilon_0 := \min\{\delta/(2C_1), \delta/(2C_2)\}$, we find that the conditions $y_n \in D(\delta/2)$ and $x_n \in D(\delta)$ are ensured for $0 \le n \le T/\varepsilon$, and it follows that $|x_n - \check{y}(\varepsilon n)| \le (C_1 + C_2)\varepsilon < \delta$ also holds for $0 \le n \le T/\varepsilon$. Finally, since x_n, y_n, and $y(n) = \check{y}(\varepsilon n)$ remain in $D(\delta)$ for $0 \le n \le T/\varepsilon$, and since f and g coincide on $D(\delta)$, we see that whenever $0 \le \varepsilon < \varepsilon_0$, the dynamics of systems ($1'$), ($2'$), and ($3'$) coincide with the dynamics of the respective systems (1.1), (1.2), and (1.3) on the interval $0 \le n \le T/\varepsilon$, which completes the proof. //

Remark 4.11 In the above proof, the order constants C_1, C_2 and the threshold ε_0 depend on δ. We note that, since the motions of systems (1.1), (1.2), and (1.3) remain in the δ-tube $D(\delta)$, the uniform norms which appear in the proofs of Theorem 1, 2, and 3 may be taken over $\overline{D}(\delta)$ rather than all of $\mathbf{R}^d$.

4.3.3 Proof of Proposition C

Let $g(u,n) := (f(x, nq/p + \tau),\, a)^{\mathrm{T}}$ where $u := (x,\tau)^{\mathrm{T}}$, so that $g : U \times \mathbf{R} \times \mathbf{N} \to \mathbf{R}^{d+1}$. The system $u_{n+1} = u_n + \varepsilon g(u_n, n)$ clearly satisfies the hypotheses of Proposition B applied to Theorem 1, with d replaced by $d+1$ and U replaced by $U \times \mathbf{R}$. Thus the conclusion of Theorem 1 applies to u_n as well as to x_n. The constants c and c' may be easily estimated along the lines of the proofs of Theorem 1 and Lemma 4 respectively. Taking into account Remark 4.11, we find $c(T,|a|) = \big(L_g T\|\widehat{g}\|_{\overline{D}} + \|\widetilde{g}\|_{\overline{D}\times\mathbf{N}}\big)e^{L_g T}$ and $c'(T,|a|) = TL_g\|\widehat{g}\|_{\overline{D}}\,e^{L_g T}$. Here L_g is the u-Lipschitz constant of g, which is independent of a. On the other hand, the norms $\|\widehat{g}\|_{\overline{D}}$ and $\|\widetilde{g}\|_{\overline{D}\times\mathbf{N}}$ depend on $|a|$, since $\|\widehat{g}\|_{\overline{D}} = \sup_{v\in\overline{D}} \sqrt{|\widehat{f}(v)|^2 + a^2}$ and $\|\widetilde{g}\|_{\overline{D}\times\mathbf{N}} = \sup_{v\in\overline{D},n\in\mathbf{N}} \sqrt{|\widehat{f}(v) - f(x, nq/p+\tau)|^2 + a^2}$. //

Appendix.

In this appendix, for the sake of completeness we supply statements and proofs of two elementary results with which the reader may be unfamiliar.

Lemma 3 (The Gronwall inequality for sequences). *Let $A \geq 0$, $B \geq 0$, and $\{E_n\}_{n=0}^{\infty}$ be a sequence of nonnegative real numbers with $E_0 = 0$ satisfying $E_N \;\leq\; A\sum_{n=0}^{N-1} E_n \;+\; B$. Then $E_N \;\leq\; Be^{AN}$.*

Proof. Set $R_{N-1} \;=\; A\sum_{n=0}^{N-1} E_n \;+\; B$ so that $R_N - R_{N-1} \;=\; AE_N \;\leq\; AR_{N-1} \;\Rightarrow\; R_N \;\leq\; (1+A)R_{N-1}$. Proceeding inductively, we find that $R_N \leq (1+A)R_{N-1} \leq \ldots \leq (1+A)^N R_0 = B(1+A)^N \leq Be^{AN}$, where we have used $R_0 = B$ and $x > 0 \Rightarrow (1+x)^{1/x} \leq e$. //

Lemma 4 (Equivalence of autonomous flows and maps). *Let $\varepsilon > 0$, and suppose $\overline{f} : \mathbf{R}^d \to \mathbf{R}^d$ is Lipschitz continuous and has compact support. Then*

the map $\quad y_{n+1} = y_n + \varepsilon\overline{f}(y_n) \quad (1.5) \qquad$ *and the flow* $\quad \dfrac{dy}{dt} = \varepsilon\overline{f}(y) \quad (1.6)$

are equivalent in the sense that there exists a constant $K > 0$ such that the solutions y_n and $y(t)$ of (1.5) and (1.6), respectively, with common initial condition $y_0 = y(0) \in \mathbf{R}^d$ satisfy the nearness condition $|y_n - y(n)| \;\leq\; K\varepsilon$ for $0 \leq n \leq T/\varepsilon$.

Proof. Let $L > 0$ denote the global Lipschitz constant of $\overline{f}$. First we note that $y(n+1) - y(n) = \varepsilon\int_n^{n+1} \overline{f}(y(t))\,dt = \varepsilon\overline{f}(y(n)) + \varepsilon\int_n^{n+1}\big(\overline{f}(y(t)) - \overline{f}(y(n))\big)\,dt$. Thus $y_{n+1} - y(n+1) = y_n - y(n) + \varepsilon\big(\overline{f}(y_n) - \overline{f}(y(n))\big) - \varepsilon\int_n^{n+1}\big(\overline{f}(y(t)) - \overline{f}(y(n))\big)\,dt$. Now setting $E_n = |y_n - y(n)|$, we obtain $E_{n+1} \leq E_n + \varepsilon L E_n + \varepsilon^2 L\|\overline{f}\|_{\mathbf{R}^d}$, since $\varepsilon\big|\int_n^{n+1}\big(\overline{f}(y(t)) - \overline{f}(y(n))\big)\,dt\big| \;\leq\; \varepsilon L\int_n^{n+1} |y(t) - y(n)|\,dt \;\leq\; \varepsilon^2 L\|\overline{f}\|_{\mathbf{R}^d}$. Using this last inequality to form a telescoping sum, we arrive to $E_n - E_0 \;\leq\; \varepsilon L\sum_{k=0}^{n-1} E_k + n\varepsilon^2 L\|\overline{f}\|_{\mathbf{R}^d}$, or $E_n \leq \varepsilon L\sum_{k=0}^{n-1} E_k + \varepsilon TL\|\overline{f}\|_{\mathbf{R}^d}$ (since $E_0 = 0$ and $0 \leq n \leq T/\varepsilon$). Finally, we apply the Gronwall inequality for sequences (Lemma 3, above) to get $E_n \leq \varepsilon TL\|\overline{f}\|_{\mathbf{R}^d}\, e^{\varepsilon Ln} \leq \varepsilon TL\|\overline{f}\|_{\mathbf{R}^d}\, e^{LT}$, so the desired conclusion is true with $K = TL\|\overline{f}\|_{\mathbf{R}^d}\, e^{LT}$. //

Acknowledgments. HSD thanks F. Golse and L. Michelotti for stimulating discussions, and acknowledges the support of the C.P. Taft Foundation at the University of Cincinnati. JAE and MV gratefully acknowledge discussions with L. Michelotti and T. Sen, as well as support from DOE grant DE-FG03-99ER41104.

References

[AMR] R. Abraham, J. Marsden, and T. Ratiu, *Manifolds, Tensor Analysis, and Applications* (2nd Ed.), Springer-Verlag, New York, 1988.

[ABG] M. Andreolli, D. Bambusi, and A. Giorgilli, On a weakened form of the averaging principle in multifrequency systems, *Nonlinearity* **8** (2): 283–293 (1995).

[Ar] V.I. Arnold, Proof of A.N. Kolmogorov's theorem on the preservation of quasi-periodic motions under small perturbations of the Hamiltonian [Russian], *Usp. Mat. Nauk. SSSR* **18** (5): 13–40 (1963) [English translation: *Russian Math. Surveys* **18** (5): 9–36 (1963)].

[BTT] A. Bazzani, E. Todesco, and G. Turchetti, A normal form approach to the theory of nonlinear betatronic motion, *CERN Report* 94–02, Geneva, 1994.

[Bel] E.P. Belan, On the averaging method in the theory of finite difference equations [Russian], *Ukrain. Mat. Z.* **19**, no. 3: 85–90 (1967).

[Bes] J. Besjes, On the asymptotic methods for non-linear differential equations, *J. Mécanique* **8**: 357–372 (1969).

[BM] N.N. Bogoliubov and Y.A. Mitropolsky, *Asymptotic Methods in the Theory of Non-Linear Oscillations* (2nd Ed.) [translated from Russian], Gordon and Breach Science Publishers, New York, 1961.

[BHS] H.W. Broer, G.B. Huitema, and M.B. Sevryuk, *Quasiperiodic Motions in Families of Dynamical Systems*, Lecture Notes in Mathematics, Vol. 1645, Springer-Verlag, New York, 1996.

[Br] A. Browder, *Mathematical Analysis*, Springer-Verlag, New York, 1996.

[Dr] V.A. Dragan, Method of averaging for systems of sum-difference equations [Russian], *Mat. Issled.* (*Computational Methods of Mechanics*), No. 64: 172–181, 195–196 (1981); Methods of averaging and freezing of systems of finite difference equations of two variables [Russian], *Mat. Issled.*, No. 64: 182–188, 196–197 (1981).

[DEG] H.S. Dumas, J.A. Ellison, and F. Golse, A mathematical theory of planar particle channeling in crystals, *Physica D* **146** (1–4): 341–366 (2000).

[DESV] H.S. Dumas, J.A. Ellison, T. Sen, and M. Vogt, work in preparation.

[DEV] H.S. Dumas, J.A. Ellison, and M. Vogt, presentation at 2002 Spring Meeting of APS, in Albuquerque, NM; see `http://www.math.unm.edu/~ellison/papers/APS02_MAP.ps.gz`

[ES] J.A. Ellison and H.-J. Shi, The method of averaging in beam dynamics, in *Accelerator Physics Lectures at the Superconducting Super Collider* (AIP Conf. Procs. 326, Y. Yan and M. Syphers, Eds.): 590–632 (1995).

[Kö] T.W. Körner, *Fourier Analysis*, Cambridge University Press, Cambridge, 1988.

[Ne] A. Neishtadt (private communication w/HSD), June, 2000.

[PW] R.E.A.C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain*, AMS Colloquium Publications, Vol. 19, New York, 1934.

[Rü] H. Rüßmann, On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus, in *Lecture Notes in Physics* **38**, 598–624, Springer-Verlag, New York, 1975; On the frequencies of quasi periodic solutions of analytic integrable Hamiltonian systems, in *Seminar on Dynamical Systems* (Proceedings of the Euler International Mathematical Institute, St. Petersburg, 1991; V. Lazutkin et al., Eds.), 160–183, Birkhäuser, Berlin, 1994.

[Sá] A.W. Sáenz, Higher order averaging for nonperiodic systems, *J. Math. Phys.* **41**: 5342–5368 (2000).

[Yo] J.-C. Yoccoz, An introduction to small divisors problems, in *From Number Theory to Physics* (Les Houches, 1989), 659–679, Springer-Verlag, Berlin, 1992.