# Hamiltonian normal forms and applications to maps

MARKUS KUNZE[1] & RAFAEL ORTEGA[2]

[1] Universität Köln, Institut für Mathematik, Weyertal 86-90,
D - 50931 Köln, Germany

[2] Departamento de Matemática Aplicada, Universidad de Granada,
E-18071 Granada, Spain

**Key words:** Hamiltonian normal form, exponential averaging

2010 *Mathematics Subject Classification.* 37J40, 37E40, 34C29, 70K65

**Abstract**

We discuss some of the results in [6] and give fully detailed proofs.

## 1 Introduction

To motivate this mainly technical note, we consider holomorphic twist maps $f:(\theta,r)\mapsto(\theta_1,r_1)$ of the form

$$\theta_1=\theta+\frac{1}{r^\alpha}(\gamma+F_1(\theta,r)),\quad r_1=r+r^{1-\alpha}F_2(\theta,r),\tag{1.1}$$

where $\alpha\in]0,1[$ and $\gamma\in\mathbb{R}\setminus\{0\}$. Such maps occur for instance in the description of the holomorphic Fermi-Ulam ping-pong [6, 4], and the variable $\theta$ is not assumed to be periodic. Rather, (1.1) is defined on a set

$$\Omega=\mathbb{R}_\delta\times\{r\in\mathbb{C}:\mathrm{Re}\,r>\underline{r},\,|\mathrm{Im}\,r|<\eta|r|\}$$

for some $\delta,\underline{r}>0$ and $\eta\in]0,1[$, where $\mathbb{R}_\delta=\{\theta\in\mathbb{C}:|\mathrm{Im}\,\theta|<\delta\}$ denotes the open strip in the complex plane about $\mathbb{R}$ of width $\delta$. In [4], the main assumptions are:

(i) the smallness of the holomorphic functions $F_j$ on $\Omega$ (supposed to map reals into reals), in the sense that $F_j(\theta,r)=\mathcal{O}(r^{-\alpha})$, uniformly in $\theta\in\mathbb{R}_\delta$, for $j=1,2$;

(ii) $\mathfrak{h}(\theta,r)=\mathfrak{h}_0(\theta,r)+\mathcal{O}(r^{1-2\alpha})$ uniformly in $\theta\in\mathbb{R}_\delta$, where $r_1\,d\theta_1-r\,d\theta=d\mathfrak{h}$ holds for (1.1), with $\mathfrak{h}_0(\theta,r)=-\frac{\alpha\gamma}{1-\alpha}\,r^{1-\alpha}$ corresponding to $F_1=F_2=0$.

In order to investigate the possible boundedness or growth of the $(r_n)_{n\in\mathbb{N}_0}$ in a forward complete real orbit $(\theta_n,r_n)_{n\in\mathbb{N}_0}$ of (1.1), it is convenient to rescale $\xi=\varepsilon^{1/\alpha}r$, which puts $f$ from (1.1) into the form $P_\varepsilon:G_\rho\to\mathbb{C}^2$, given by

$$P_\varepsilon:\quad x_1=x+\varepsilon l(x,\varepsilon),\quad x_1=(q_1,p_1),\quad x=(q,p).\tag{1.2}$$

Here $x_1 = (q_1, p_1) = (\theta_1, \xi_1)$, $x = (q, p) = (\theta, \xi)$ and $l(x, \varepsilon) = (\frac{1}{\xi^\alpha}(\gamma + F_1(\theta, \frac{\xi}{\varepsilon^{1/\alpha}})), \xi^{1-\alpha} F_2(\theta, \frac{\xi}{\varepsilon^{1/\alpha}}))$. It turns out that, for $\varepsilon > 0$ small enough, the family of maps $\{P_\varepsilon\}$ can be defined on a common domain $G_\rho$, where $G = \mathbb{R} \times ]1, 2[$ and

$$G_\rho = \{x = (q, p) \in \mathbb{C}^2 : |\operatorname{Im} q| < \rho,\ \operatorname{dist}(p, I) < \rho\}.$$

This leads us to study general maps $P_\varepsilon : G_\rho \to \mathbb{C}^2$ of the form (1.2), where $l$ belongs to a certain class of maps $\mathcal{M}_{1,\rho,\sigma}$ that has to be carefully set up in order to account for singularities of $l$ or $\frac{\partial l}{\partial \varepsilon}$ at $\varepsilon = 0$; recall the definition of $l$ in the ping-pong example.

Inspired by [3], we call the family of maps $\{P_\varepsilon\}$ E-symplectic, if $p_1\, dq_1 - p\, dq = dh(\cdot, \varepsilon)$ for a function $h \in \mathcal{M}_{1,\rho,\sigma}$ such that, as $\varepsilon \to 0$,

$$h(q, p, \varepsilon) = \varepsilon\, \mathfrak{m}(q, p) + \mathcal{O}(\varepsilon^2), \quad \frac{\partial h}{\partial \varepsilon}(q, p, \varepsilon) = \mathfrak{m}(q, p) + \mathcal{O}(\varepsilon),$$

uniformly in $(q, p) \in G_\rho$ for a bounded function $\mathfrak{m} : G_\rho \to \mathbb{C}$. It turns out (see [4]) that all these conditions can be verified for the ping-pong after rescaling $\mathfrak{h}$ from (ii) to $h$. Furthermore, it is possible to construct a function $E = E(x)$ satisfying $J \nabla E(x) = l(x, 0)$, where $J$ denotes the standard symplectic matrix. The function $E$ should be thought of as an approximate first integral (adiabatic invariant) for the family $\{P_\varepsilon\}$. This means that the variation of $E$ along the orbit remains small for an exponentially long time. More precisely, we have the following result, which should be compared to [6, (2.7), p. 135 and Prop. 3, p. 136].

**Theorem 1.1** *Suppose that $l \in \mathcal{M}_{1,\rho,\sigma}$, and for $\varepsilon \in [0, \sigma]$ consider the family of maps $P_\varepsilon : G_\rho \to \mathbb{C}^2$ given by*

$$P_\varepsilon : \quad x_1 = x + \varepsilon l(x, \varepsilon). \tag{1.3}$$

*Let the family $\{P_\varepsilon\}$ be E-symplectic. Then there exist $\hat{\sigma} \in ]0, \sigma]$ and constants $\hat{C}, \hat{D} > 0$ (depending upon $\rho$, $\sigma$, $\|l\|_{1,\rho,\sigma}$, the interval $I$, $\|h\|_{1,\rho,\sigma}$ and $\sup_{\varepsilon \in ]0,\sigma]} \|\varepsilon^{-1}(\frac{\partial h}{\partial \varepsilon}(\cdot, \varepsilon) - \mathfrak{m})\|_\rho$) such that if*

$$(x_n)_{0 \le n \le N} = (P_\varepsilon^n(x_0))_{0 \le n \le N}$$

*is a real forward orbit piece of $P_\varepsilon$ so that $x_n \in G$ for all $0 \le n \le N$, then*

$$|E(x_n) - E(x_0)| \le \hat{C}\varepsilon, \quad 0 \le n \le \min\{N, N_\varepsilon\}, \quad N_\varepsilon = [e^{\hat{D}/\varepsilon}]. \tag{1.4}$$

It is the purpose of this note to present a fully detailed proof of Theorem 1.1, along the lines that are indicated in [6]. It is based on realizing $P_\varepsilon$ as the Poincaré map of a periodic Hamiltonian system. One main difficulty is that the class $\mathcal{M}_{1,\rho,\sigma}$ has to allow for a non-smooth dependence of $l$ on $\varepsilon$, since this is what is needed in the applications. Furthermore, the fact that in $G = \mathbb{R} \times ]1, 2[$ the first coordinate can be unbounded poses some technical challenges; this is accounted for by introducing assumptions on the primitive of the 1-form that are not explicit in [6]. Therefore we need to introduce suitable function classes $\mathcal{H}_{\rho,\sigma}$ and $\widetilde{\mathcal{H}}_{\rho,\sigma}$ for the relevant Hamiltonians $H = H(x, t, \varepsilon)$ in the Hamiltonian normal form theorem (see Section 3).

## 2 E-symplectic families of maps

An important observation in [6] is the existence of adiabatic invariants for families of analytic canonical maps close to the identity. Given a convex domain $G \subset \mathbb{R}^N \times \mathbb{R}^N$ and a family of symplectic maps

$$P_\varepsilon : G \to \mathbb{R}^N \times \mathbb{R}^N, \quad x_1 = x + \varepsilon l(x, \varepsilon),$$

it is possible to construct a function $E = E(x)$ satisfying

$$J\nabla E(x) = l(x, 0), \tag{2.1}$$

where $J = \begin{pmatrix} 0 & I_N \\ -I_N & 0 \end{pmatrix}$. For small $\varepsilon$ the iteration $x_{n+1} = P_\varepsilon(x_n)$ can be interpreted as a numerical integration method for the Hamiltonian system $\dot{x} = J\nabla E(x)$. This fact suggests that $E(x)$ should be an adiabatic invariant for $P_\varepsilon$, meaning that

$$|E(P_\varepsilon^n(x)) - E(x)| \le C\varepsilon, \quad 0 \le n \le N_\varepsilon, \tag{2.2}$$

where $N_\varepsilon$ is of the order $e^{D/\varepsilon}$; the constants $C, D > 0$ should only depend upon an appropriate norm of $l$. In essence this is discussed in Remark 5 and Proposition 3 of [6]. Additional details can be found in [2], in particular in the case of bounded domains.

However, the previous statements must be taken with some caution in the case where the underlying domain is unbounded. As a counter-example we consider the family of translations

$$x_1 = x + \varepsilon Jv + \varepsilon^2 v,$$

defined on the whole space $G = \mathbb{R}^N \times \mathbb{R}^N$. Here $v \neq 0$ is a fixed vector and $E(x) = \langle x, v \rangle$ satisfies (2.1), since $l(x, \varepsilon) = Jv + \varepsilon v$. Due to $P_\varepsilon^n(x) = x + n\varepsilon l(x, \varepsilon)$ we obtain

$$|E(P_\varepsilon^n(x)) - E(x)| = \varepsilon^2 n |v|^2.$$

Therefore (2.2) does hold only for $n \le N_\varepsilon = \mathcal{O}(1/\varepsilon)$ many steps.

To overcome this inherent difficulty, Benettin and Giorgilli in [2] considered an unbounded domain $G$ and a family of maps derived from a symplectic integration algorithm for a Newtonian system of the type $\ddot{q} = -\nabla V(q)$. Then they impose some growth conditions on $V(q)$ as $|q| \to \infty$. We will follow a different approach and assume that our family $\{P_\varepsilon\}$ satisfies a condition inspired by the notion of an exact symplectic map (called E-symplectic), as it was understood in our previous work [3]. Furthermore, to simplify matters, we will restrict ourselves to the case of direct interest to us for applications. Throughout we will take

$$N = 1 \quad \text{and} \quad G = \mathbb{R} \times I,$$

where $I \subset \mathbb{R}$ is an open and bounded interval. Our goal will be to understand the dynamics of a map on the plane $(\theta, r) \mapsto (\theta_1, r_1)$ when $r \to \infty$. For this reason our family of maps $\{P_\varepsilon\}$, $P_\varepsilon : (q, p) \mapsto (q_1, p_1)$, will be obtained after a rescaling $q = \theta$, $p = \varepsilon r$ with $q \in \mathbb{R}$ and $p \in ]1, 2[$. This procedure will lead to functions $l(x, \varepsilon)$ that are analytic in $x$, but not necessarily smooth in $\varepsilon$; a prototype can be the function $l(x, \varepsilon) = h(x/\varepsilon^2)$, where $h$ is real analytic in $[1, \infty[$ and

$h(\zeta) \to 0$ as $\zeta \to \infty$. Then $l$ is continuous as a function of the two variables $(x, \varepsilon)$, but the partial derivatives $\partial_\varepsilon^k l$ do not always exist at $\varepsilon = 0$.

The following definitions are motivated by the previous discussions. In general, for the norms on $\mathbb{C}^d$ and $\mathbb{C}^{d_1 \times d_2}$ we will take $|x| = \max_{1 \le i \le d} |x_i|$ and $|A| = \max_{1 \le i \le d_1, 1 \le j \le d_2} |a_{ij}|$, respectively. Note that for $A \in \mathbb{C}^{d \times d}$, $x \in \mathbb{C}^d$, $A_1 \in \mathbb{C}^{d_1 \times d}$ and $A_2 \in \mathbb{C}^{d \times d_2}$ this implies

$$|Ax| \le d|A||x|, \quad |A_1 A_2| \le d|A_1||A_2|.$$

The points in $G = \mathbb{R} \times I$ will be denoted by $x = (q, p)$. For $\rho > 0$ we will write

$$G_\rho = \{x = (q, p) \in \mathbb{C}^2 : |\mathrm{Im}\, q| < \rho,\ \mathrm{dist}(p, I) < \rho\}.$$

Given $\varphi : G_\rho \to \mathbb{C}$ holomorphic, let

$$\|\varphi\|_\rho = \sup\{|\varphi(x)| : x \in G_\rho\}.$$

If $0 < r < \rho$, then by the Cauchy integral formula one has

$$\|D\varphi\|_r \le \frac{1}{\rho - r} \|\varphi\|_\rho,$$

where $D\varphi$ is the Jacobian.

**Definition 2.1 (The classes $\mathcal{M}_{\rho,\sigma}$ and $\mathcal{M}_{1,\rho,\sigma}$)** *Let $\rho > 0$ and $\sigma \in ]0, 1[$.*

*(i) The class $\mathcal{M}_{\rho,\sigma}$ consists of those continuous maps $l : G_\rho \times [0, \sigma] \to \mathbb{C}^2$, $l = l(x, \varepsilon)$, which satisfy:*

*(a) $l$ maps real into reals; and*

*(b) for every $\varepsilon \in [0, \sigma]$ the map $l(\cdot, \varepsilon)$ is holomorphic on $G_\rho$ and*

$$\|l\|_{\rho,\sigma} = \sup\{\|l(\cdot, \varepsilon)\|_\rho : \varepsilon \in [0, \sigma]\} < \infty.$$

*(ii) The class $\mathcal{M}_{1,\rho,\sigma}$ consists of those continuous maps $l : G_\rho \times [0, \sigma] \to \mathbb{C}^2$, $l = l(x, \varepsilon)$, satisfying*

*(a) $l$ maps real into reals;*

*(b) $l$ is $C^\infty$ in $G_\rho \times ]0, \sigma]$;*

*(c) for every $\varepsilon \in [0, \sigma]$ the map $l(\cdot, \varepsilon)$ is holomorphic on $G_\rho$;*

*(d) one has*

$$\|l\|_{1,\rho,\sigma} = \|l\|_{\rho,\sigma} + \sup\left\{\left\|\frac{\partial l}{\partial \varepsilon}(\cdot, \varepsilon)\right\|_\rho : \varepsilon \in ]0, \sigma]\right\} < \infty.$$

**Remark 2.2** Note that, for a map $l \in \mathcal{M}_{\rho,\sigma}$ or $l \in \mathcal{M}_{1,\rho,\sigma}$, all the derivatives $\partial_x^\alpha \partial_\varepsilon^k l(\cdot, \varepsilon) : G_\rho \to \mathbb{C}^2$ for $\varepsilon \in ]0, \sigma]$ are holomorphic, where $\alpha \in \mathbb{N}_0^2$ and $k \in \mathbb{N}_0$. Similarly, all the $\partial_x^\alpha l : G_\rho \times [0, \sigma] \to \mathbb{C}^2$ are continuous functions of both variables. This follows from the Cauchy integral formula and the continuity of $l$. Furthermore, the derivatives can be interchanged: $\partial_x^\alpha \partial_\varepsilon^k l(\cdot, \varepsilon) = \partial_\varepsilon^k \partial_x^\alpha l(\cdot, \varepsilon)$.

**Definition 2.3** *Suppose that $l \in \mathcal{M}_{1,\rho,\sigma}$, and for $\varepsilon \in [0,\sigma]$ consider the family of maps $P_\varepsilon : G_\rho \to \mathbb{C}^2$ given by*

$$P_\varepsilon : \quad x_1 = x + \varepsilon l(x,\varepsilon), \quad x_1 = (q_1, p_1), \quad x = (q,p). \tag{2.3}$$

*We say that the family $\{P_\varepsilon\}$ is E-symplectic, if there is a function $h \in \mathcal{M}_{1,\rho,\sigma}$ such that*

$$p_1\, dq_1 - p\, dq = dh(\cdot, \varepsilon) \tag{2.4}$$

*and there exists a bounded function $\mathfrak{m} : G_\rho \to \mathbb{C}$ satisfying*

$$h(q,p,\varepsilon) = \varepsilon\, \mathfrak{m}(q,p) + \mathcal{O}(\varepsilon^2) \quad \text{as} \quad \varepsilon \to 0 \tag{2.5}$$

*and*

$$\frac{\partial h}{\partial \varepsilon}(q,p,\varepsilon) = \mathfrak{m}(q,p) + \mathcal{O}(\varepsilon) \quad \text{as} \quad \varepsilon \to 0 \tag{2.6}$$

*uniformly in $(q,p) \in G_\rho$.*

**Remark 2.4** (a) $\mathfrak{m}$ is holomorphic in $G_\rho$. To see this, note that $\frac{\partial h}{\partial \varepsilon}(\cdot, \varepsilon)$ is holomorphic for $\varepsilon > 0$ by Remark 2.2. Since $\mathfrak{m}$ is the uniform limit of $\int_0^1 \frac{\partial h}{\partial \varepsilon}(q,p,t\varepsilon)\, dt$ as $\varepsilon \to 0$, it is holomorphic itself.
(b) $\mathfrak{m}$ satisfies

$$\frac{\partial \mathfrak{m}}{\partial q}(q,p) = p\, \frac{\partial l_1}{\partial q}(q,p,0) + l_2(q,p,0), \quad \frac{\partial \mathfrak{m}}{\partial p}(q,p) = p\, \frac{\partial l_1}{\partial p}(q,p,0), \tag{2.7}$$

where $l = (l_1, l_2)$. For, we observe from (2.5) that $\varepsilon^{-1} h \to \mathfrak{m}$ uniformly on $G_\rho$. Therefore also the derivatives converge, uniformly on compact subsets of $G_\rho$. From (2.4),

$$\varepsilon^{-1} \frac{\partial h}{\partial q} = l_2 + p\, \frac{\partial l_1}{\partial q} + \varepsilon\, l_2\, \frac{\partial l_1}{\partial q}, \quad \varepsilon^{-1} \frac{\partial h}{\partial p} = p\, \frac{\partial l_1}{\partial p} + \varepsilon\, l_2\, \frac{\partial l_1}{\partial p}.$$

Thus it remains to pass to the limit $\varepsilon \to 0$ and use Remark 2.2. Relation (2.7) can also be stated as

$$\nabla \mathfrak{m}(x) = p\, \nabla l_1(x,0) + \begin{pmatrix} l_2(x,0) \\ 0 \end{pmatrix}, \quad x = (q,p). \tag{2.8}$$

(c) One has

$$\frac{\partial l_1}{\partial q}(q,p,0) + \frac{\partial l_2}{\partial p}(q,p,0) = 0, \tag{2.9}$$

as follows from $\frac{\partial^2 \mathfrak{m}}{\partial q \partial p} = \frac{\partial^2 \mathfrak{m}}{\partial p \partial q}$. Relation (2.9) implies that the Jacobian matrix $Dl(x,0)$ is Hamiltonian, i.e., it satisfies $Dl(x,0)^* J + J Dl(x,0) = 0$, or equivalently, $J Dl(x,0)$ is symmetric. Since $G_\rho$ is simply connected, we conclude that there is a holomorphic function $E : G_\rho \to \mathbb{C}$ such that $J \nabla E = l(\cdot, 0)$, i.e., (2.1) holds. Actually, (2.8) shows that we can take

$$E(x) = l_1(x,0)p - \mathfrak{m}(x), \quad x = (q,p). \tag{2.10}$$

(d) The relation $J\nabla E = l(\cdot, 0)$ yields

$$dE = \frac{\partial E}{\partial q}\, dq + \frac{\partial E}{\partial p}\, dp = -l_2\, dq + l_1\, dp.$$

Hence $E(x) = E(x_0) + \int_\gamma (-l_2\, dq + l_1\, dp)$ for every path $\gamma$ that connects a fixed $x_0 \in G$ to $x$. This observation makes the connection to the formula for $E$ given in [6] below (2.7).
(e) Condition (2.6) does not follow from (2.5), as the example

$$h(q, p, \varepsilon) = \varepsilon\, \mathfrak{m}(q, p) + \varepsilon^2 \sin\left(\frac{1}{\varepsilon}\right)$$

shows.

# 3 A Hamiltonian normal form

In this section we will give fully detailed proofs of some of the results in [6] and we will discuss the assumptions that are needed for those proofs to work.

**Definition 3.1 (The class $\mathcal{H}_{\rho,\sigma}$)** *For $\rho > 0$ and $\sigma \in ]0, 1[$ let $\mathcal{H}_{\rho,\sigma}$ be the class of continuous functions $H : G_\rho \times \mathbb{R} \times [0, \sigma] \to \mathbb{C}$, $H = H(x, t, \varepsilon)$, satisfying*

*(a) $H$ is $T$-periodic in $t$;*

*(b) $H$ maps reals into reals;*

*(c) for every $t \in \mathbb{R}$ and $\varepsilon \in [0, \sigma]$ the function $H(\cdot, t, \varepsilon)$ is holomorphic on $G_\rho$; and*

*(d) the gradient w.r. to $x$, $\nabla H = \nabla_x H(q, p, t, \varepsilon)$, is a continuous function from $G_\rho \times \mathbb{R} \times [0, \sigma]$ to $\mathbb{C}^2$ such that*

$$\|\nabla H\|_{\rho,\sigma} := \sup\left\{\|\nabla H(\cdot, \cdot, t, \varepsilon)\|_\rho : t \in \mathbb{R}, \varepsilon \in [0, \sigma]\right\} < \infty.$$

**Remark 3.2** Note that for a function $H \in \mathcal{H}_{\rho,\sigma}$ all partial derivatives $\partial_x^\alpha H : G_\rho \times \mathbb{R} \times [0, \sigma] \to \mathbb{C}^d$ w.r. to $x$ are again continuous functions of all three variables, where as usual $\partial_x^\alpha H = \frac{\partial^{|\alpha|}}{\partial_q^{\alpha_1} \partial_p^{\alpha_2}} H$ for a multi-index $\alpha \in \mathbb{N}_0^2$. This is a consequence of the fact that the Cauchy integral formula can be differentiated w.r. to $x$.

**Definition 3.3 (The class $\widetilde{\mathcal{H}}_{\rho,\sigma}$)** *The class $\widetilde{\mathcal{H}}_{\rho,\sigma}$ consists of those $H \in \mathcal{H}_{\rho,\sigma}$ with the additional property that*

$$\int_0^T H(x, t, \varepsilon)\, dt = 0 \tag{3.1}$$

*for $x \in G_\rho$ and $\varepsilon \in [0, \sigma]$.*

Observe that if $H \in \widetilde{\mathcal{H}}_{\rho,\sigma}$, then $t \mapsto \int_0^t H(x,s,\varepsilon)\,ds$ is $T$-periodic.

For $h \in \widetilde{\mathcal{H}}_{\rho,\sigma}$ consider the (time-dependent) implicit Euler transformation $\Phi : (x,t,\varepsilon) \mapsto y$ with inverse $(y,t,\varepsilon) \mapsto x = \Psi(y,t,\varepsilon)$, $x = (q,p)$, $y = (q_1,p_1)$, which is given by

$$q_1 = q - \varepsilon \int_0^t \frac{\partial h}{\partial p_1}(q,p_1,s,\varepsilon)\,ds, \quad p_1 = p + \varepsilon \int_0^t \frac{\partial h}{\partial q}(q,p_1,s,\varepsilon)\,ds. \tag{3.2}$$

Solving the second equation, we obtain $p_1 = p_1(q,p,t,\varepsilon)$, and the first equation then determines $q_1 = q_1(q,p,t,\varepsilon)$. We will show that the map $\Psi$ is well-defined and it is an admissible change of variables, in a sense that is made precise in the following definition.

**Definition 3.4** *Let $0 < \rho_1 \le \rho$ and $0 < \sigma_1 \le \sigma$. A map $\Psi : G_{\rho_1} \times \mathbb{R} \times [0,\sigma_1] \to \mathbb{C}^2$, $x = \Psi(y,t,\varepsilon)$, will be called an admissible change of variables, if it satisfies*

- *(a) $\Psi$ maps reals into reals;*
- *(b) $\Psi$ is $T$-periodic in $t$ and $\Psi(y,0,\varepsilon) = \Psi(y,T,\varepsilon) = y$;*
- *(c) $\Psi$ is continuous;*
- *(d) for every $t \in \mathbb{R}$ and $\varepsilon \in [0,\sigma_1]$ the map $\Psi(\cdot,t,\varepsilon)$ is holomorphic in $G_{\rho_1}$, and for every $y \in G_{\rho_1}$ and $\varepsilon \in [0,\sigma_1]$ the map $\Psi(y,\cdot,\varepsilon) \in C^1(\mathbb{R})$;*
- *(e) all admissible partial derivatives with regard to $y$ and $t$ are continuous functions of all the arguments $(y,t,\varepsilon)$; and*
- *(f) for every $t \in \mathbb{R}$ and $\varepsilon \in [0,\sigma_1]$ the map $\Psi(\cdot,t,\varepsilon)$ is a symplectic diffeomorphism from $G_{\rho_1}$ onto its image.*

**Lemma 3.5** *For $0 < r < \rho$ and $\sigma > 0$ given, let $\sigma_1 = \min\{\frac{\rho-r}{12},\sigma\}$. Then, for each $h \in \widetilde{\mathcal{H}}_{\rho,\sigma}$ with $T\|\nabla h\|_{\rho,\sigma} \le 1$, the equations (3.2) define a map*

$$\Psi : G_r \times \mathbb{R} \times [0,\sigma_1] \to \mathbb{C}^2$$

*that is an admissible change of variables and satisfies $\Psi(G_r,t,\varepsilon) \subset G_\rho$ for every $t \in \mathbb{R}$ and $\varepsilon \in [0,\sigma_1]$. Moreover,*

$$\|\Psi(\cdot,\cdot,\varepsilon) - I\|_r \le \varepsilon T \|\nabla h\|_{\rho,\sigma} \tag{3.3}$$

*for $\varepsilon \in [0,\sigma_1]$.*

**Remarks 3.6** (a) The simple geometry of $G$ implies the following useful fact: if $(q,p),(q_1,p_1) \in G_\rho$, then also $(q,p_1),(q_1,p) \in G_\rho$. For this reason the equations (3.2) are well-defined.
(b) The condition $T\|\nabla h\|_{\rho,\sigma} \le 1$ is just imposed to get a definitive value for $\sigma_1$. When we are going to apply the lemma to an arbitrary $h \in \widetilde{\mathcal{H}}_{\rho,\sigma}$ later, a rescaling argument (in $\varepsilon$) can be used.

**Proof of Lemma 3.5:** To solve the first equation in (3.2) we evenly split the interval $[r, \rho]$ into $r < R_1 < R_2 < \rho$, where $R_1 - r = R_2 - R_1 = \rho - R_2 = \frac{\rho - r}{3}$. Define

$$X = \{q \in \mathbb{C} : |\mathrm{Im}\, q| \le R_2\}$$

as well as $\hat{\sigma} = \min\{\frac{\rho - r}{6}, \sigma\}$. For $(q_1, p_1) \in G_{R_1}$ and $\varepsilon \in [0, \hat{\sigma}]$ fixed, let

$$\mathcal{F}(q) = q_1 + \varepsilon \int_0^t \frac{\partial h}{\partial p_1}(q, p_1, s, \varepsilon)\, ds.$$

Then $\mathcal{F} : X \to X$ is a self-map, since $\varepsilon T \|\nabla h\|_{\rho,\sigma} \le \varepsilon \le \hat{\sigma} < \frac{\rho - r}{3} = R_2 - R_1$. The condition (3.1) in Definition 3.3 allows us to restrict to the time interval $[0, T]$. From the Cauchy integral formula we deduce

$$\|D^2 h\|_{R_2,\sigma} \le \frac{1}{\rho - R_2} \|\nabla h\|_{\rho,\sigma} = \frac{3}{\rho - r} \|\nabla h\|_{\rho,\sigma}.$$

This estimate applies in particular to the cross-derivative $\frac{\partial^2 h}{\partial q \partial p_1}$ and ensures that $\mathcal{F}$ is a contraction, due to

$$\varepsilon T \|D^2 h\|_{R_2,\sigma} \le \varepsilon T \frac{3}{\rho - r} \|\nabla h\|_{\rho,\sigma} \le \frac{3\hat{\sigma}}{\rho - r} \le \frac{1}{2}. \tag{3.4}$$

The unique fixed point of $\mathcal{F}$ defines a continuous map $q = q(q_1, p_1, t, \varepsilon) : G_{R_1} \times \mathbb{R} \times [0, \hat{\sigma}] \to X$ Then the definition of $\Psi$ is completed by setting

$$p = p_1 - \varepsilon \int_0^t \frac{\partial h}{\partial q}(q(q_1, p_1, t, \varepsilon), p_1, s, \varepsilon)\, ds.$$

Note that

$$\mathrm{dist}(p, I) \le \mathrm{dist}(p_1, I) + \varepsilon T \|\nabla h\|_{\rho,\sigma} \le R_1 + \varepsilon \le R_1 + \hat{\sigma} < R_2,$$

and hence $\Psi$ is defined on $G_{R_1} \times \mathbb{R} \times [0, \hat{\sigma}]$ and takes values in $\overline{G_{R_2}} \subset G_\rho$. The bound

$$|\Psi(q_1, p_1, t, \varepsilon) - (q_1, p_1)| \le \varepsilon T \|\nabla h\|_{\rho,\sigma} \tag{3.5}$$

is a direct consequence of the definition of $\Psi$, and in particular (3.5) implies (3.3), since $r < R_1$.

To prove the smoothness of $\Psi$, we observe that $q$ is defined implicitly by the equation $F = 0$, where

$$F(q, q_1, p_1, t, \varepsilon) = q - q_1 - \varepsilon \int_0^t \frac{\partial h}{\partial p_1}(q, p_1, s, \varepsilon)\, ds.$$

The transversality condition

$$\frac{\partial F}{\partial q} = 1 - \varepsilon \int_0^t \frac{\partial^2 h}{\partial q \partial p_1}(q, p_1, s, \varepsilon)\, ds \ne 0$$

is satisfied, due to (3.4). Hence the implicit function theorem applies to yield that $q$ (and hence $p$) verifies all the smoothness requirements for an admissible change of variables.

It remains to establish that $\Psi(\cdot, t, \varepsilon)$ is a symplectic diffeomorphism from $G_r$ onto its image, for $t \in [0, T]$ and $\varepsilon \in [0, \sigma_1]$. Using (3.5), which is valid for $y = (q_1, p_1) \in G_{R_1}$, we deduce that

$$|D\Psi(q_1, p_1, t, \varepsilon) - I| \le \frac{\varepsilon}{R_1 - r} T \|\nabla h\|_{\rho,\sigma} = \frac{3\varepsilon}{\rho - r} T \|\nabla h\|_{\rho,\sigma} \le \frac{3\sigma_1}{\rho - r} \le \frac{1}{4}, \tag{3.6}$$

where $D\Psi = D_y\Psi = D_{(q_1,p_1)}\Psi$ is the Jacobian. This will allow us to interpret $\Psi(\cdot,t,\varepsilon)$ as a Lipschitz continuous perturbation of the identity. Indeed, if we define $\Gamma = \Psi - I$, then owing to the convexity of $G_r$ and from (3.6) we obtain the bound

$$\begin{aligned}|\Gamma(y,t,\varepsilon) - \Gamma(\tilde{y},t,\varepsilon)| &= \Big|\int_0^1 \frac{d}{ds}\left[\Gamma(sy + (1-s)\tilde{y},t,\varepsilon)\right] ds\Big| \\ &\le 2\cdot\frac{1}{4}\,|y-\tilde{y}| = \frac{1}{2}\,|y-\tilde{y}|\end{aligned}$$

for $y,\tilde{y} \in G_r$. Hence the Lipschitz constant of $\Gamma(\cdot,t,\varepsilon)$ is $\le 1/2$. This in turn implies that $\Psi(\cdot,t,\varepsilon)$ is one-to-one on $G_r$. According to (3.6), i.e., $|D\Psi(q_1,p_1,t,\varepsilon) - I| \le 1/4$, the matrix $D\Psi$ has an inverse. Thus the inverse function theorem can be applied at each fixed $y \in G_r$ to deduce that $\Psi(\cdot,t,\varepsilon)$ is a diffeomorphism from $G_r$ onto the open set $\Psi(G_r,t,\varepsilon) \subset G_\rho$. This diffeomorphism is symplectic, because it has been obtained from the equations (3.2), which can be derived from the generating function

$$S(q,p_1,t,\varepsilon) = qp_1 - \varepsilon\int_0^t h(q,p_1,s,\varepsilon)\,ds. \tag{3.7}$$

This completes the proof of the lemma. □

**Corollary 3.7** *Under the assumptions of Lemma 3.5, let* $0 < \hat{r} < r < \rho$ *and denote by* $\Psi : y \mapsto x$ *the map that is induced by (3.2). Let* $\sigma_2 = \min\{\sigma_1, \frac{r-\hat{r}}{2}\} = \min\{\frac{\rho-r}{12}, \frac{r-\hat{r}}{2}, \sigma\}$*. If* $t \in \mathbb{R}$ *and* $\varepsilon \in [0,\sigma_2]$*, then* $\Psi(G_r,t,\varepsilon) \supset G_{\hat{r}}$*.*

For the proof, the following result will be helpful, which is [**?**, Prop. I.3, p. 50].

**Lemma 3.8** *Let* $X, Y$ *be Banach spaces and suppose that* $U \subset Y$ *is open. If* $\Psi : U \to \Psi(U) \subset X$ *is a homeomorphism,* $\Psi^{-1}$ *is Lipschitz continuous with constant* $\mathrm{Lip}(\Psi^{-1}) < \lambda$*, and* $\overline{B_r(y)} \subset U$*, then*

$$\Psi(\overline{B_r(y)}) \supset \overline{B_{r/\lambda}(\Psi(y))}.$$

**Proof of Corollary 3.7:** We are going to apply Lemma 3.8 with $U = G_r$, $\Psi = \Psi(\cdot,t,\varepsilon)$ and $\lambda = 2$. Inspecting the proof of Lemma 3.5, we have shown in (3.6) that $|D\Psi(y) - I| \le 1/4$ for $y \in G_{R_1} \supset G_r$, where we write $\Psi(y) = \Psi(y,t,\varepsilon)$. This yields

$$|D\Psi(y)^{-1}| = \Big|\sum_{j=0}^{\infty}(-1)^j (D\Psi(y) - I)^j\Big| \le \sum_{j=0}^{\infty} 2^j\,|D\Psi(y) - I|^j \le 2.$$

Next observe that if $y \in G_{\hat{r}}$, then $\overline{B_{r-\hat{r}}(y)} \subset G_r$, as a consequence of the geometry of $G$ and the choice of the norm. For $y \in G_{\hat{r}}$ we also have $|y - \Psi(y)| \le \varepsilon$ by (3.3), which means that $y \in \overline{B_\varepsilon(\Psi(y))}$. Owing to Lemma 3.8 we obtain

$$y \in \overline{B_\varepsilon(\Psi(y))} \subset \overline{B_{(r-\hat{r})/2}(\Psi(y))} \subset \Psi(\overline{B_{r-\hat{r}}(y)}) \subset \Psi(G_r),$$

as claimed. □

**Lemma 3.9** *For $0 < r < \rho$ and $\sigma > 0$ given, let $H \in \mathcal{H}_{\rho,\sigma}$ and $h \in \widetilde{\mathcal{H}}_{\rho,\sigma}$ be such that $T\|\nabla h\|_{\rho,\sigma} \leq 1$. Defining $\sigma_1 = \min\{\frac{\rho-r}{12}, \sigma\}$ as before, we consider the admissible change of variables $x = \Psi(y,t,\varepsilon)$ for $(y,t,\varepsilon) \in G_r \times \mathbb{R} \times [0,\sigma_1]$ according to Lemma 3.5. Then, for every $\varepsilon \in [0,\sigma_1]$, the $T$-periodic Hamiltonian system*

$$\dot{x} = \varepsilon J \nabla_x H(x,t,\varepsilon) \tag{3.8}$$

*is transformed (pulled back via $\Psi$) into*

$$\dot{y} = \varepsilon J \nabla_y K(y,t,\varepsilon), \tag{3.9}$$

*where*

$$K(y,t,\varepsilon) = H(\Psi(y,t,\varepsilon),t,\varepsilon) - h(q(y,t,\varepsilon),p_1,t,\varepsilon); \tag{3.10}$$

*recall that $y = (q_1,p_1)$, $\Psi = (\Psi_1,\Psi_2)$ with $\Psi_1 = q$ and $\Psi_2 = p$. Moreover, $K \in \mathcal{H}_{r,\sigma_1}$ and*

$$\|\nabla K\|_{r,\sigma_1} \leq 3\,\|\nabla H\|_{\rho,\sigma} + \frac{5}{2}\,\|\nabla h\|_{\rho,\sigma}. \tag{3.11}$$

**Proof:** Given a Hamiltonian system $\dot{x} = J\nabla_x H(x,t)$ and a change of variables $x = \Psi(y,t)$ that is induced by a generating function of the type $S = S(q,p_1,t)$, the pull-back of the system is $\dot{y} = J\nabla_y K(y,t)$, where

$$K(y,t) = H(\Psi(y,t),t) + \frac{\partial S}{\partial t}(q(y,t),p_1,t).$$

This is part of the classical theory of non-autonomous Hamiltonian systems, cf. [1]. It was known early on, see [7, pp. 13-16] for an elegant exposition. In our case a generating function $S$ of $\Psi$ is given in (3.7), and the formula (3.10) follows.

To show that $K \in \mathcal{H}_{r,\sigma_1}$ we differentiate (3.10) to obtain

$$\nabla K = (D\Psi)^* \nabla H - \frac{\partial h}{\partial q}\,\nabla q - \frac{\partial h}{\partial p_1}\,(0,1)^*.$$

From (3.6) we know that $|D\Psi(y) - I| \leq 1/4$ for $y \in G_{R_1} \supset G_r$, which in turn yields $|D\Psi(y)| \leq 5/4$. In particular, also $|\nabla q| \leq 5/4$, dropping the arguments. Therefore

$$|\nabla K| \leq 2|D\Psi||\nabla H| + \left|\frac{\partial h}{\partial q}\right| |\nabla q| + \left|\frac{\partial h}{\partial p_1}\right| \leq \frac{5}{2}\,|\nabla H| + \frac{5}{4}\left|\frac{\partial h}{\partial q}\right| + \left|\frac{\partial h}{\partial p_1}\right|$$

leads to (3.11). □

Given $H \in \mathcal{H}_{\rho,\sigma}$ we define the function

$$\bar{H}(x,\varepsilon) = \frac{1}{T}\int_0^T H(x,t,\varepsilon)\,dt \quad \text{and} \quad \tilde{H} = H - \bar{H}.$$

Then $\bar{H} \in \mathcal{H}_{\rho,\sigma}$ is autonomous and $\tilde{H} \in \widetilde{\mathcal{H}}_{\rho,\sigma}$. Moreover, we have the bounds

$$\|\nabla \bar{H}\|_{\rho,\sigma} \leq \|\nabla H\|_{\rho,\sigma} \quad \text{and} \quad \|\nabla \tilde{H}\|_{\rho,\sigma} \leq 2\|\nabla H\|_{\rho,\sigma}. \tag{3.12}$$

**Lemma 3.10** *For $0 < r < \rho$ and $\sigma > 0$ given, let $H \in \mathcal{H}_{\rho,\sigma}$ be such that $T\|\nabla \tilde{H}\|_{\rho,\sigma} \le 1$. We apply Lemma 3.9 with $h = \tilde{H}$. Then the admissible change of variables $\Psi : y \mapsto x$ and the new Hamiltonian function $K$ satisfy*

$$\begin{aligned} \|\Psi(\cdot,\cdot,\varepsilon) - I\|_r &\le \varepsilon T \|\nabla \tilde{H}\|_{\rho,\sigma}, && (3.13)\\ \|\bar{K} - \bar{H}\|_{r,\sigma_1} &\le 2\varepsilon T\, \|\nabla \tilde{H}\|_{\rho,\sigma}\, (\|\nabla \bar{H}\|_{\rho,\sigma} + \|\nabla \tilde{H}\|_{\rho,\sigma}), && (3.14)\\ \|\tilde{K}\|_{r,\sigma_1} &\le 4\varepsilon T\, \|\nabla \tilde{H}\|_{\rho,\sigma}\, (\|\nabla \bar{H}\|_{\rho,\sigma} + \|\nabla \tilde{H}\|_{\rho,\sigma}), && (3.15) \end{aligned}$$

*for $\varepsilon \in [0, \sigma_1]$.*

**Proof:** The first estimate (3.13) is a direct consequence of (3.3). To derive (3.14), we rewrite (3.10) in the form

$$K(y,t,\varepsilon) - \bar{H}(y,\varepsilon) = \bar{H}(\Psi(y,t,\varepsilon),\varepsilon) - \bar{H}(y,\varepsilon) + \tilde{H}(\Psi(y,t,\varepsilon),t,\varepsilon) - \tilde{H}(q(y,t,\varepsilon),p_1,t,\varepsilon). \quad (3.16)$$

Since $q = \Psi_1$ is just a coordinate of $\Psi$,

$$|\Psi(y,t,\varepsilon) - (q(y,t,\varepsilon),p_1)| = |p(y,t,\varepsilon) - p_1| \le |\Psi(y,t,\varepsilon) - y|,$$

and hence it follows from (3.16) and (3.13) that

$$\begin{aligned} \|K - \bar{H}\|_{r,\sigma_1} &\le 2\|\nabla \bar{H}\|_{\rho,\sigma}\, \|\Psi - I\|_{r,\sigma_1} + 2\|\nabla \tilde{H}\|_{\rho,\sigma}\, \|\Psi - I\|_{r,\sigma_1} \\ &\le 2\varepsilon T \|\nabla \tilde{H}\|_{\rho,\sigma} (\|\nabla \bar{H}\|_{\rho,\sigma} + \|\nabla \tilde{H}\|_{\rho,\sigma}). && (3.17) \end{aligned}$$

Observing that

$$\begin{aligned} \|\bar{K} - \bar{H}\|_{r,\sigma_1} &= \sup\left\{ \left| \frac{1}{T}\int_0^T [K(y,t,\varepsilon) - \bar{H}(y,\varepsilon)]\, dt \right| : y \in G_r, \varepsilon \in [0,\sigma_1] \right\} \\ &\le \sup\{ |K(y,t,\varepsilon) - \bar{H}(y,\varepsilon)| : y \in G_r, t \in \mathbb{R}, \varepsilon \in [0,\sigma_1]\} \\ &= \|K - \bar{H}\|_{r,\sigma_1}, && (3.18) \end{aligned}$$

(3.14) is a consequence of (3.17). Concerning (3.15), it suffices to write $\tilde{K} = (K - \bar{H}) + (\bar{H} - \bar{K})$ and to use (3.18) as well as (3.17). □

For the next result we are going to apply Lemma 3.10 $N$ times.

**Lemma 3.11** *Let $0 < r < \rho$ and $\sigma > 0$ be given. For every integer $N \ge 1$ and $H \in \mathcal{H}_{\rho,\sigma}$ so that $T\|\nabla H\|_{\rho,\sigma} \le 1/2$ there exists an admissible change of variables $x = \Psi_N(y,t,\varepsilon)$, which is defined on $G_r \times \mathbb{R} \times [0, \sigma_N]$ for*

$$\sigma_N = \min\left\{ \frac{\rho - r}{72N}, \sigma \right\},$$

*and which satisfies $\Psi(G_r, t, \varepsilon) \subset G_\rho$ for $t \in \mathbb{R}$ and $\varepsilon \in [0, \sigma_N]$. Furthermore,*

$$\dot{x} = \varepsilon J \nabla_x H(x,t,\varepsilon) \quad (3.19)$$

*is transformed (pulled back via $\Psi_N$) into*

$$\dot{y} = \varepsilon J \nabla_y H_N(y,t,\varepsilon) \tag{3.20}$$

*for $H_N \in \mathcal{H}_{r,\sigma_N}$, and moreover we have*

$$\begin{aligned} \|\Psi_N(\cdot,\cdot,\varepsilon) - I\|_r &\leq 2\varepsilon, & (3.21)\\ \|\nabla \tilde{H}_N\|_{r,\sigma_N} &\leq \left(\frac{1}{T}\right) 2^{-N}, & (3.22)\\ \|\nabla \bar{H}_N\|_{r,\sigma_N} &\leq \frac{3}{2T}, & (3.23)\\ |\bar{H}_N(y,\varepsilon) - \bar{H}_N(y,0)| &\leq \frac{24}{T}\,\varepsilon + |\bar{H}(y,\varepsilon) - \bar{H}(y,0)|, & (3.24)\end{aligned}$$

*for $y \in G_r$ and $\varepsilon \in [0,\sigma_N]$.*

First we are going to state an auxiliary result that will be useful in the proof of this lemma.

**Lemma 3.12** *Let $(b_k)_{0\leq k\leq K}$ and $(c_k)_{0\leq k\leq K}$ for some $K \in \mathbb{N}\cup\{\infty\}$ be sequences of positive numbers such that*

$$b_k \leq \alpha b_{k-1}(b_{k-1}+c_{k-1}) \quad \textit{and} \quad c_k \leq b_k + c_{k-1}$$

*for $1 \leq k \leq K$, where $\alpha > 0$ is such that $4\alpha(b_0+c_0) \leq 1$. Then*

$$b_k \leq \frac{1}{2^k}\, b_0 \quad \textit{and} \quad c_k \leq b_0 + c_0$$

*for $0 \leq k \leq K$.*

**Proof:** We check that $b_k \leq 2^{-k} b_0$ and $c_k \leq b_0 \sum_{j=1}^k 2^{-j} + c_0$ by induction. Clearly this holds for $k=0$. For the induction step, by hypothesis we have

$$b_{k+1} \leq \alpha b_k(b_k+c_k) \leq \alpha b_0 2^{-k}(b_0 2^{-k} + b_0 + c_0) \leq 2\alpha b_0 2^{-k}(b_0+c_0) \leq b_0 2^{-(k+1)},$$

and hence in particular

$$c_{k+1} \leq b_{k+1} + c_k \leq b_0 2^{-(k+1)} + b_0 \sum_{j=1}^{k} 2^{-j} + c_0 = b_0 \sum_{j=1}^{k+1} 2^{-j} + c_0,$$

which completes the argument. □

**Proof of Lemma 3.11:** We introduce a uniform partition of the interval $[r,\rho]$ by

$$\rho_N = r < \rho_{N-1} < \ldots < \rho_1 < \rho_0 = \rho,$$

where $\rho_k - \rho_{k+1} = \frac{\rho - r}{N}$ for $k = 0,\ldots,N-1$. The midpoint of $[\rho_{k+1},\rho_k]$ will be denoted by $r_{k+1}$, so that $\rho_k - r_{k+1} = r_{k+1} - \rho_{k+1} = \frac{\rho-r}{2N}$.

Set $H_0 = H$ and observe that $T\|\nabla \tilde{H}_0\|_{\rho,\sigma} = T\|\nabla \tilde{H}\|_{\rho,\sigma} \le 2T\|\nabla H\|_{\rho,\sigma} \le 1$ by (3.12) and by assumption. Hence we can apply Lemma 3.10 for $r$ replaced by $r_1$ to obtain an admissible change of variables $\Psi^{(1)}$ that is defined on $G_{r_1} \times \mathbb{R} \times [0, \hat{\sigma}_1]$ and takes values in $G_\rho$, where $\hat{\sigma}_1 = \min\{\frac{\rho-r_1}{12}, \sigma\} = \min\{\frac{\rho-r}{24N}, \sigma\}$. The transformed Hamiltonian is denoted by $H_1 \in \mathcal{H}_{r_1,\hat{\sigma}_1}$, and from (3.13)–(3.15) we have the bounds

$$\begin{aligned}
\|\Psi^{(1)}(\cdot,\cdot,\varepsilon) - I\|_{r_1} &\le \varepsilon T \|\nabla \tilde{H}_0\|_{\rho,\sigma}, & (3.25)\\
\|\bar{H}_1 - \bar{H}_0\|_{r_1,\hat{\sigma}_1} &\le 2\varepsilon T\, \|\nabla \tilde{H}_0\|_{\rho,\sigma}\, (\|\nabla \bar{H}_0\|_{\rho,\sigma} + \|\nabla \tilde{H}_0\|_{\rho,\sigma}), & (3.26)\\
\|\tilde{H}_1\|_{r_1,\hat{\sigma}_1} &\le 4\varepsilon T\, \|\nabla \tilde{H}_0\|_{\rho,\sigma}\, (\|\nabla \bar{H}_0\|_{\rho,\sigma} + \|\nabla \tilde{H}_0\|_{\rho,\sigma}), & (3.27)
\end{aligned}$$

for $\varepsilon \in [0, \hat{\sigma}_1]$. Since $\sigma_N \le \hat{\sigma}_1$, we may replace $\hat{\sigma}_1$ by $\sigma_N$ in all of the above. Next we are going to derive some preliminary estimates on $H_0$ and $H_1$. Let

$$b_0 = \|\nabla \tilde{H}_0\|_{\rho,\sigma} \quad \text{and} \quad c_0 = \|\nabla \bar{H}_0\|_{\rho,\sigma}$$

as well as

$$b_1 = \|\nabla \bar{H}_1 - \nabla \bar{H}_0\|_{\rho_1,\sigma_N} + \|\nabla \tilde{H}_1\|_{\rho_1,\sigma_N} \quad \text{and} \quad c_1 = \|\nabla \bar{H}_1\|_{\rho_1,\sigma_N}.$$

Note that by (3.12),

$$b_0 \le 2\|\nabla H\|_{\rho,\sigma} \le \frac{1}{T} \quad \text{and} \quad c_0 \le \|\nabla H\|_{\rho,\sigma} \le \frac{1}{2T}. \tag{3.28}$$

Furthermore,

$$b_k \le \frac{12N\sigma_N T}{\rho - r}\, b_{k-1}(b_{k-1} + c_{k-1}) \quad \text{and} \quad c_k \le b_k + c_{k-1} \tag{3.29}$$

are verified for $k = 1$. To establish this claim, note that by the Cauchy integral formula, (3.26) and (3.27),

$$\begin{aligned}
b_1 &= \|\nabla \bar{H}_1 - \nabla \bar{H}_0\|_{\rho_1,\sigma_N} + \|\nabla \tilde{H}_1\|_{\rho_1,\sigma_N} \le \frac{1}{r_1 - \rho_1} \Big( \|\bar{H}_1 - \bar{H}_0\|_{r_1,\sigma_N} + \|\tilde{H}_1\|_{r_1,\sigma_N} \Big)\\
&\le \frac{2N}{\rho - r}\, (2\sigma_N T + 4\sigma_N T)\, \|\nabla \tilde{H}_0\|_{\rho,\sigma}\, (\|\nabla \bar{H}_0\|_{\rho,\sigma} + \|\nabla \tilde{H}_0\|_{\rho,\sigma}) = \frac{12N\sigma_N T}{\rho - r}\, b_0(b_0 + c_0).
\end{aligned} \tag{3.30}$$

Concerning the bound on $c_1$, we have

$$c_1 = \|\nabla \bar{H}_1\|_{\rho_1,\sigma_N} \le \|\nabla \bar{H}_1 - \nabla \bar{H}_0\|_{\rho_1,\sigma_N} + \|\nabla \bar{H}_0\|_{\rho_1,\sigma_N} \le b_1 + c_0. \tag{3.31}$$

We are going to prove that this process can be repeated $N$ times, if we consider the sequence of nested domains

$$G_r = G_{\rho_N} \subset G_{r_N} \subset G_{\rho_{N-1}} \subset \ldots \subset G_{r_2} \subset G_{\rho_1} \subset G_{r_1} \subset G_{\rho_0} = G_\rho.$$

We will find a sequence $\Psi^{(k)}$, $k = 1, \ldots, N$, of admissible changes of variables sending the set $G_{r_k} \times \mathbb{R} \times [0, \sigma_N]$ into $G_{\rho_{k-1}}$. These changes of variable $\Psi^{(k)}$ and Hamiltonian functions $H_k \in \mathcal{H}_{r_k,\sigma_N} \subset \mathcal{H}_{\rho_k,\sigma_N}$ will be constructed by finite induction w.r.t. to $k \in \{1, \ldots, N\}$. Suppose that $\Psi^{(1)}, \dots, \Psi^{(k)}$ and $H_1, \dots, H_k$ have already been obtained, with the additional property that (3.29) holds, where

$$b_k = \|\nabla \bar{H}_k - \nabla \bar{H}_{k-1}\|_{\rho_k,\sigma_N} + \|\nabla \tilde{H}_k\|_{\rho_k,\sigma_N} \quad \text{and} \quad c_k = \|\nabla \bar{H}_k\|_{\rho_k,\sigma_N}$$

for $k \geq 1$. With $\alpha = \frac{12N\sigma_N T}{\rho - r}$ we note that

$$4\alpha(b_0 + c_0) \leq 1,$$

since by (3.12) and our hypotheses

$$\frac{48N\sigma_N T}{\rho - r}\left(\|\nabla \tilde{H}_0\|_{\rho,\sigma} + \|\nabla \bar{H}_0\|_{\rho,\sigma}\right) \leq \frac{144N\sigma_N T}{\rho - r}\|\nabla H\|_{\rho,\sigma} \leq \frac{72N\sigma_N}{\rho - r} \leq 1.$$

Hence Lemma 3.12 applies to yield $b_k \leq 2^{-k} b_0$ and $c_k \leq b_0 + c_0$. In particular, it follows from (3.28) that

$$T\|\nabla \tilde{H}_k\|_{\rho_k,\sigma_N} \leq T b_k \leq T b_0 \leq 1,$$

and Lemma 3.10 is applicable, for $r$ replaced by $r_{k+1}$ and $\sigma$ replaced by $\sigma_N$. The resulting admissible change of variables $\Psi^{(k+1)}$ is defined on $G_{r_{k+1}} \times \mathbb{R} \times [0, \hat{\sigma}_{k+1}]$ and takes values in $G_{\rho_k}$, where $\hat{\sigma}_{k+1} = \min\{\frac{\rho_k - r_{k+1}}{12}, \sigma_N\} = \min\{\frac{\rho - r}{24N}, \sigma_N\} = \sigma_N$. The transformed Hamiltonian is denoted by $H_{k+1} \in \mathcal{H}_{r_{k+1},\sigma_N}$, and from (3.13)–(3.15) we deduce the bounds

$$\begin{aligned}
\|\Psi^{(k+1)}(\cdot,\cdot,\varepsilon) - I\|_{r_{k+1}} &\leq \varepsilon T \|\nabla \tilde{H}_k\|_{\rho_k,\sigma_N}, && (3.32)\\
\|\bar{H}_{k+1} - \bar{H}_k\|_{r_{k+1},\sigma_N} &\leq 2\varepsilon T\, \|\nabla \tilde{H}_k\|_{\rho_k,\sigma_N} (\|\nabla \bar{H}_k\|_{\rho_k,\sigma_N} + \|\nabla \tilde{H}_k\|_{\rho_k,\sigma_N}), && (3.33)\\
\|\tilde{H}_{k+1}\|_{r_{k+1},\sigma_N} &\leq 4\varepsilon T\, \|\nabla \tilde{H}_k\|_{\rho_k,\sigma_N} (\|\nabla \bar{H}_k\|_{\rho_k,\sigma_N} + \|\nabla \tilde{H}_k\|_{\rho_k,\sigma_N}), && (3.34)
\end{aligned}$$

for $\varepsilon \in [0, \sigma_N]$. Analogously to (3.30) and (3.31), it follows from the Cauchy integral formula in conjunction with (3.33) and (3.34) that (3.29) holds for $k+1$. Therefore the inductive process to obtain the $\Psi^{(k)}$ and $H_k$ can be completed up to $k = N$.

For the estimate (3.22), note that by (3.29) and (3.28)

$$\|\nabla \tilde{H}_N\|_{r,\sigma_N} = \|\nabla \tilde{H}_N\|_{\rho_N,\sigma_N} \leq b_N \leq 2^{-N} b_0 \leq \frac{1}{T}\, 2^{-N}.$$

The bound (3.23) is also a consequence of (3.29) and (3.28), since

$$\|\nabla \bar{H}_N\|_{r,\sigma_N} = \|\nabla \bar{H}_N\|_{\rho_N,\sigma_N} = c_N \leq b_0 + c_0 \leq \frac{3}{2T}.$$

The desired admissible change of variables $\Psi_N$ is defined as the composition

$$\Psi_N = \Psi^{(1)} \circ \Psi^{(2)} \circ \dots \circ \Psi^{(N)},$$

which is defined on $G_r \times \mathbb{R} \times [0, \sigma_N]$ and takes values in $G_\rho$. To obtain (3.21), we are going to use the formula

$$\Psi_N(\cdot,\cdot,\varepsilon) - I = \sum_{k=1}^{N-1} [(\Psi^{(k)} - I) \circ \Psi^{(k+1)} \circ \dots \circ \Psi^{(N)}](\cdot,\cdot,\varepsilon) + (\Psi^{(N)}(\cdot,\cdot,\varepsilon) - I). \qquad (3.35)$$

For $k \geq 1$ the composition $\Psi^{(k+1)} \circ \ldots \circ \Psi^{(N)}$ maps $G_r$ into $G_{\rho_k} \subset G_{r_k}$ in $y$. Therefore due to (3.32), with $k+1$ replaced by $k$, and using (3.28),

$$\begin{aligned} \|[(\Psi^{(k)} - I) \circ \Psi^{(k+1)} \circ \ldots \circ \Psi^{(N)}](\cdot,\cdot,\varepsilon)\|_r &\leq \|\Psi^{(k)}(\cdot,\cdot,\varepsilon) - I\|_{r_k} \leq \varepsilon T \|\nabla \tilde{H}_{k-1}\|_{\rho_{k-1},\sigma_N} \\ &\leq \varepsilon T b_{k-1} \leq \varepsilon T 2^{-(k-1)} b_0 \leq \varepsilon 2^{-(k-1)}. \end{aligned}$$

Analogously,

$$\|\Psi^{(N)}(\cdot,\cdot,\varepsilon) - I\|_r = \|\Psi^{(N)}(\cdot,\cdot,\varepsilon) - I\|_{\rho_N} \leq \|\Psi^{(N)}(\cdot,\cdot,\varepsilon) - I\|_{r_N} \leq \varepsilon 2^{-(N-1)}.$$

Using (3.35), the foregoing estimates in turn lead to

$$\begin{aligned} \|\Psi_N(\cdot,\cdot,\varepsilon) - I\|_r &\leq \sum_{k=1}^{N-1} \|[(\Psi^{(k)} - I) \circ \Psi^{(k+1)} \circ \ldots \circ \Psi^{(N)}](\cdot,\cdot,\varepsilon)\|_r + \|\Psi^{(N)}(\cdot,\cdot,\varepsilon) - I\|_r \\ &\leq \sum_{k=1}^{N-1} \varepsilon 2^{-(k-1)} + \varepsilon 2^{-(N-1)} \leq 2\varepsilon, \end{aligned}$$

which is (3.21). To prove (3.24), we first note that by (3.33) and (3.28),

$$\begin{aligned} \|\bar{H}_{k+1} - \bar{H}_k\|_{r,\sigma_N} &\leq \|\bar{H}_{k+1} - \bar{H}_k\|_{r_{k+1},\sigma_N} \\ &\leq 2\varepsilon T \,\|\nabla \tilde{H}_k\|_{\rho_k,\sigma_N} \,(\|\nabla \bar{H}_k\|_{\rho_k,\sigma_N} + \|\nabla \tilde{H}_k\|_{\rho_k,\sigma_N}) \\ &\leq 2\varepsilon T b_k (b_k + c_k) \\ &\leq 2\varepsilon T 2^{-k} b_0 (2^{-k} b_0 + b_0 + c_0) \\ &\leq 2^{-k} \left(\frac{6}{T}\right) \varepsilon. \end{aligned}$$

For $y \in G_r$ and $\varepsilon \in [0, \sigma_N]$ it hence follows that

$$\begin{aligned} &|\bar{H}_N(y,\varepsilon) - \bar{H}_N(y,0)| \\ &= \Big| \sum_{k=0}^{N-1} (\bar{H}_{k+1}(y,\varepsilon) - \bar{H}_k(y,\varepsilon)) - \sum_{k=0}^{N-1} (\bar{H}_{k+1}(y,0) - \bar{H}_k(y,0)) + (\bar{H}_0(y,\varepsilon) - \bar{H}_0(y,0)) \Big| \\ &\leq 2 \sum_{k=0}^{N-1} \|\bar{H}_{k+1} - \bar{H}_k\|_{r,\sigma_N} + |\bar{H}(y,\varepsilon) - \bar{H}(y,0)| \\ &\leq \frac{12}{T} \varepsilon \sum_{k=0}^{N-1} 2^{-k} + |\bar{H}(y,\varepsilon) - \bar{H}(y,0)| \leq \frac{24}{T} \varepsilon + |\bar{H}(y,\varepsilon) - \bar{H}(y,0)|. \end{aligned}$$

This completes the proof of Lemma 3.11. □

Now we are in a position to derive the “Hamiltonian normal form” with exponentially small remainder. For our particular domain $G = \mathbb{R} \times I$, this is essentially the result that is announced in [6, Remark 2, p. 134]. To prepare for the statement, we need to introduce a more relaxed class of transformations, as compared to Definition 3.4.

**Definition 3.13** *Let $0 < \rho_1 \le \rho$ and $0 < \sigma_1 \le \sigma$. A map $\Psi : G_{\rho_1} \times \mathbb{R} \times [0, \sigma_1] \to \mathbb{C}^2$, $x = \Psi(y, t, \varepsilon)$, will be called a change of variables, if it satisfies*

*(a) $\Psi$ maps reals into reals;*

*(b) $\Psi$ is $T$-periodic in $t$ and $\Psi(y, 0, \varepsilon) = \Psi(y, T, \varepsilon) = y$;*

*(c) for every $\varepsilon \in [0, \sigma_1]$ the map $\Psi(\cdot, \cdot, \varepsilon)$ is $C^1$ in the real sense, and for every $t \in \mathbb{R}$ and $\varepsilon \in [0, \sigma_1]$ the map $\Psi(\cdot, t, \varepsilon)$ is holomorphic in $G_{\rho_1}$; and*

*(d) for every $t \in \mathbb{R}$ and $\varepsilon \in [0, \sigma_1]$ the map $\Psi(\cdot, t, \varepsilon)$ is a symplectic diffeomorphism from $G_{\rho_1}$ onto its image.*

Note that we are not assuming any property of continuous dependence w.r. to the parameter $\varepsilon$. This is in contrast to the previous notion of an admissible change of variables, introduced in Definition 3.4.

**Theorem 3.14** *For $0 < r < \rho$ and $\sigma > 0$ given, let $H \in \mathcal{H}_{\rho,\sigma}$. Then there exist $C, D > 0$ (depending upon $T, r, \rho, \|\nabla H\|_{\rho,\sigma}$) with the following properties. There is a change of variables $x = \Psi(y, t, \varepsilon)$, which is defined on $G_r \times \mathbb{R} \times [0, \sigma]$ and which satisfies $\Psi(G_r, t, \varepsilon) \subset G_\rho$ for $t \in \mathbb{R}$ and $\varepsilon \in [0, \sigma]$, such that*

$$\dot{x} = \varepsilon J \nabla_x H(x, t, \varepsilon) \tag{3.36}$$

*is transformed (pulled back via $\Psi$) into*

$$\dot{y} = \varepsilon (J \nabla_y \mathcal{N}(y, \varepsilon) + J \nabla_y \mathcal{R}(y, t, \varepsilon)), \tag{3.37}$$

*for functions $\mathcal{N} \in \mathcal{H}_{r,\sigma}$ and $\mathcal{R} \in \widetilde{\mathcal{H}}_{r,\sigma}$. Furthermore,*

$$\|\Psi(\cdot, \cdot, \varepsilon) - I\|_r \le C\varepsilon, \tag{3.38}$$

$$\|\nabla_y \mathcal{N}(\cdot, \varepsilon)\|_r \le C, \tag{3.39}$$

$$\|\nabla_y \mathcal{R}(\cdot, \cdot, \varepsilon)\|_r \le C e^{-D/\varepsilon}, \tag{3.40}$$

$$|\mathcal{N}(y, \varepsilon) - \mathcal{N}(y, 0)| \le C\varepsilon + \|\bar{H}(\cdot, \varepsilon) - \bar{H}(\cdot, 0)\|_\rho, \tag{3.41}$$

*for $y \in G_r$ and $\varepsilon \in [0, \sigma]$. In addition,*

$$\mathcal{N}(y, 0) = \bar{H}(y, 0). \tag{3.42}$$

**Proof:** We are going to show that

$$C = \max\left\{2\lambda, \frac{2\lambda}{T}, \frac{24\lambda^2}{T}\right\} \quad \text{and} \quad D = \frac{\rho - r}{144\lambda}$$

have the asserted properties, where $\lambda = 2T\|\nabla H\|_{\rho,\sigma}$. The cases $H = 0$ or $\varepsilon = 0$ are trivial, so in particular we may assume that $\lambda > 0$. We rewrite (3.36) as $\dot{x} = \hat{\varepsilon} J \nabla_x \hat{H}(x, t, \hat{\varepsilon})$, where

$\hat\varepsilon = \lambda\varepsilon \in [0,\hat\sigma]$ for $\hat\sigma = \lambda\sigma$ and $\hat H(x,t,\hat\varepsilon) = \lambda^{-1}H(x,t,\lambda^{-1}\hat\varepsilon)$. It follows that $\hat H \in \mathcal{H}_{\rho,\hat\sigma}$ satisfies $2T\|\nabla\hat H\|_{\rho,\hat\sigma} = 2T\lambda^{-1}\|\nabla H\|_{\rho,\sigma} = 1$. Thus we may apply Lemma 3.11 to $\hat H$ and with

$$N = \Big[\frac{\rho - r}{72\lambda\varepsilon}\Big].$$

Hence there exists an admissible change of variables $x = \hat\Psi(y,t,\hat\varepsilon)$, which is defined on $G_r \times \mathbb{R} \times [0,\hat\sigma_N]$ for

$$\hat\sigma_N = \min\Big\{\frac{\rho - r}{72N}, \hat\sigma\Big\},$$

and which satisfies $\hat\Psi(G_r,t,\varepsilon) \subset G_\rho$ for $t \in \mathbb{R}$ and $\varepsilon \in [0,\hat\sigma_N]$. Furthermore, $\dot x = \hat\varepsilon J\nabla_x\hat H(x,t,\hat\varepsilon)$ is transformed into $\dot y = \hat\varepsilon J\nabla_y K(y,t,\hat\varepsilon)$ for $K \in \mathcal{H}_{r,\hat\sigma_N}$, and in addition we have

$$\|\hat\Psi(\cdot,\cdot,\hat\varepsilon) - I\|_r \le 2\hat\varepsilon, \quad \|\nabla\tilde K\|_{r,\hat\sigma_N} \le \Big(\frac{1}{T}\Big)2^{-N}, \quad \|\nabla\bar K\|_{r,\hat\sigma_N} \le \frac{3}{2T},$$
$$|\bar K(y,\hat\varepsilon) - \bar K(y,0)| \le \frac{24}{T}\hat\varepsilon + |\overline{\hat H}(y,\hat\varepsilon) - \overline{\hat H}(y,0)|,$$

for $y \in G_r$ and $\hat\varepsilon \in ]0,\hat\sigma_N]$. Define

$$\Psi(y,t,\varepsilon) = \hat\Psi(y,t,\lambda\varepsilon), \quad \mathcal{N}(y,\varepsilon) = \lambda\bar K(y,\lambda\varepsilon) \quad \text{and} \quad \mathcal{R}(y,t,\varepsilon) = \lambda\tilde K(y,t,\lambda\varepsilon)$$

for $y \in G_r$, $t \in \mathbb{R}$ and $\varepsilon \in ]0,\sigma]$. We also put $\Psi = I$ for $\varepsilon = 0$. If $\varepsilon \in ]0,\sigma]$, then $\hat\varepsilon = \lambda\varepsilon \le \lambda\sigma = \hat\sigma$ and moreover

$$\hat\varepsilon = \frac{1}{N}\lambda\varepsilon N \le \frac{1}{N}\lambda\varepsilon\Big(\frac{\rho - r}{72\lambda\varepsilon}\Big) = \frac{\rho - r}{72N},$$

so that $\hat\varepsilon \in ]0,\sigma_N]$. Accordingly, the first few claims are straightforwardly verified; this includes (3.38), (3.39) and (3.41). Concerning (3.40), we use the above estimate on $\nabla\tilde K$ to get for $\varepsilon \in ]0,\sigma]$

$$\begin{aligned}\|\nabla_y\mathcal{R}(\cdot,\cdot,\varepsilon)\|_r &= \lambda\|\nabla\tilde K(\cdot,\cdot,\hat\varepsilon)\|_r \le \Big(\frac{\lambda}{T}\Big)2^{-N} = \Big(\frac{2\lambda}{T}\Big)2^{-(N+1)} \le \Big(\frac{2\lambda}{T}\Big)2^{-\frac{\rho-r}{72\lambda\varepsilon}} \\ &= \Big(\frac{2\lambda}{T}\Big)4^{-\frac{\rho-r}{144\lambda\varepsilon}} \le \Big(\frac{2\lambda}{T}\Big)e^{-\frac{\rho-r}{144\lambda\varepsilon}},\end{aligned}$$

which completes the proof of (3.38)–(3.41).

Finally, with regard to (3.42), we observe that in all the previous lemmas we have $\Psi = I$ for $\varepsilon = 0$. Then we can define $\mathcal{N}(y,0) = \bar H(y,0)$, since $\bar H_k(\cdot,0) = \bar H(\cdot,0)$ for each $k$ throughout the iteration. □

**Corollary 3.15** *Under the assumptions of Theorem 3.14 let $0 < \hat r < r < \rho$ and denote by $\Psi : y \mapsto x$ the change of variables that has been constructed there. Let $\sigma_* = \min\{\frac{r-\hat r}{12C}, \sigma\}$. If $t \in \mathbb{R}$ and $\varepsilon \in [0,\sigma_*]$, then $\Psi(G_r,t,\varepsilon) \supset G_{\hat r}$.*

**Proof:** The argument is similar to the one for Corollary 3.7. □

# 4 Application to maps

Here we will prove Theorem 1.1, following the approach outlined in [6]. First we realize the map $P_\varepsilon$ as the Poincaré map of a periodic Hamiltonian system and then we are going to apply the previous results from Section 3; see [8, p. 13/14] for general information and additional references in a more abstract context.

We start with an auxiliary result on the construction of a Hamiltonian function from an exact symplectic isotopy.

**Lemma 4.1** *Assume that $\Phi : G \times [0,1] \to \mathbb{R}^2$ is $C^\infty$ and that $\Phi(\cdot,t) : G \to G(t) = \Phi(G,t)$ is a diffeomorphism for every $t \in [0,1]$. The inverse map is denoted by $\Psi(\cdot,t)$ and we will also write*

$$X = \Phi(x,t), \quad x = \Psi(X,t), \quad x = (q,p), \quad X = (Q,P), \quad \Phi = (\mathcal{F},\mathcal{G}).$$

*Assume that*

$$P\,dQ - p\,dq = d\eta(\cdot,t) \tag{4.1}$$

*for a $C^\infty$-function $\eta : G \times [0,1] \to \mathbb{R}$. Then*

$$J\nabla h_{\mathrm{aux}}(X,t) = \frac{\partial \Phi}{\partial t}(\Psi(X,t),t), \tag{4.2}$$

*where*

$$h_{\mathrm{aux}}(X,t) = \frac{\partial \mathcal{F}}{\partial t}(\Psi(X,t),t)\,\mathcal{G}(\Psi(X,t),t) - \frac{\partial \eta}{\partial t}(\Psi(X,t),t) \tag{4.3}$$

*is defined on*

$$\mathcal{D} = \{(X,t) : t \in [0,1], X \in G(t)\}.$$

**Remark 4.2** (a) Note that $G(t) \subset \mathbb{R}^2$ is open and $\mathcal{D}$ is diffeomorphic to $G \times [0,1]$ via the map $(x,t) \mapsto (\Phi(x,t),t)$. Moreover, $X(t) = \Phi(x,t)$ is a solution to $\dot{X}(t) = J\nabla h_{\mathrm{aux}}(X(t),t)$.
(b) Lemma 4.1 remains valid, if $\Phi$ and $\eta$ are $C^1$, and the cross-derivatives

$$\frac{\partial^2 \Phi}{\partial t\,\partial x} = \frac{\partial^2 \Phi}{\partial x\,\partial t}, \quad \frac{\partial^2 \eta}{\partial t\,\partial x} = \frac{\partial^2 \eta}{\partial x\,\partial t},$$

exist, coincide and are continuous functions of $(x,t)$.
(c) If $\Phi(\cdot,t)$, $\Psi(\cdot,t)$ and $\eta(\cdot,t)$ have holomorphic extensions, then also the identity (4.2) can be extended.
(d) We refer to [5, Thm. 6.2.1] for a similar result.

**Proof of Lemma 4.1 :** The identity (4.1) holds in the space of one-forms on $G$. Differentiating w.r. to $t$, we obtain

$$\frac{\partial \mathcal{G}}{\partial t}\,d\mathcal{F} + \mathcal{G}\,d\Big(\frac{\partial \mathcal{F}}{\partial t}\Big) = d\Big(\frac{\partial \eta}{\partial t}\Big).$$

It follows that

$$d\Big(\frac{\partial \mathcal{F}}{\partial t}\,\mathcal{G} - \frac{\partial \eta}{\partial t}\Big) = \frac{\partial \mathcal{F}}{\partial t}\,d\mathcal{G} + d\Big(\frac{\partial \mathcal{F}}{\partial t}\Big)\,\mathcal{G} - d\Big(\frac{\partial \eta}{\partial t}\Big) = \frac{\partial \mathcal{F}}{\partial t}\,d\mathcal{G} - \frac{\partial \mathcal{G}}{\partial t}\,d\mathcal{F} \tag{4.4}$$

on $G$. To pull back this identity under the map $\Psi(\cdot,t) : G(t) \ni X \mapsto x \in G$, denote $\mathfrak{h}(x,t) = \frac{\partial \mathcal{F}}{\partial t}(x,t)\,\mathcal{G}(x,t) - \frac{\partial \eta}{\partial t}(x,t)$. From (4.4) we thus deduce

$$\begin{aligned} dh_{\mathrm{aux}}(\cdot,t) &= d(\mathfrak{h}\circ\Psi) = d(\Psi^*\mathfrak{h}) = \Psi^*(d\mathfrak{h}) = \Psi^*\Big(\frac{\partial \mathcal{F}}{\partial t}\,d\mathcal{G} - \frac{\partial \mathcal{G}}{\partial t}\,d\mathcal{F}\Big) \\ &= \Big(\frac{\partial \mathcal{F}}{\partial t}\circ\Psi\Big)dP - \Big(\frac{\partial \mathcal{G}}{\partial t}\circ\Psi\Big)dQ, \end{aligned}$$

which is equivalent to (4.2). □

**Lemma 4.3** *Let $G = \mathbb{R}\times I \subset \mathbb{R}^2$ for an open and bounded interval $I \subset \mathbb{R}$. Suppose that $l \in \mathcal{M}_{1,\rho,\sigma}$, and for $\varepsilon \in [0,\sigma]$ consider the family of maps $P_\varepsilon : G_\rho \to \mathbb{C}^2$ given by*

$$P_\varepsilon : \quad x_1 = x + \varepsilon l(x,\varepsilon). \tag{4.5}$$

*Let the family $\{P_\varepsilon\}$ be E-symplectic and fix $0 < r < \hat r < \rho$. Then there exist $\hat\sigma \in ]0,\sigma[$ and a Hamiltonian $H_{\mathrm{aux}} \in \mathcal{H}_{\hat r,\hat\sigma}$ such that for $\varepsilon \in [0,\hat\sigma]$ the Poincaré map (time-1-map) of $\dot x = \varepsilon J\nabla H_{\mathrm{aux}}(x,t,\varepsilon)$ is $P_\varepsilon$, restricted to $G_r$. Furthermore, there exists a constant $C_{\mathrm{aux}} > 0$ such that*

$$|H_{\mathrm{aux}}(x,t,\varepsilon) - H_{\mathrm{aux}}(x,t,0)| \le C_{\mathrm{aux}}\,\varepsilon \tag{4.6}$$

*for $x \in G_{\hat r}$, $t \in [0,1]$ and $\varepsilon \in [0,\hat\sigma]$. The constant $C_{\mathrm{aux}}$ will depend upon $\rho$, $\sigma$, $r$, $\hat r$, $\|l\|_{1,\rho,\sigma}$, the interval $I$, $\|h\|_{1,\rho,\sigma}$ and $\sup_{\varepsilon\in]0,\sigma]} \|\varepsilon^{-1}(\frac{\partial h}{\partial \varepsilon}(\cdot,\varepsilon) - \mathfrak{m})\|_\rho$ (cf. the notion of E-symplecticity, Definition 2.3).*

**Proof:** Let $\chi : [0,1] \to [0,1]$ be a strictly increasing $C^\infty$-function such that $\chi(0) = 0$, $\chi(1) = 1$ and $\dot\chi(0) = \dot\chi(1) = 0$. Define

$$\Phi(x,t,\varepsilon) = x + \varepsilon\chi(t)\,l(x,\varepsilon\chi(t)) \tag{4.7}$$

and

$$\eta(x,t,\varepsilon) = h(x,\varepsilon\chi(t)).$$

For fixed $\varepsilon$ we intend to apply the relaxed version of Lemma 4.1, as outlined in Remark 4.2(b), (c). The condition (4.1) holds, due to (2.4) in Definition 2.3.

Our first aim will be to construct the inverse $\Psi$. Define $r_1 = \frac{1}{2}(\rho + \hat r)$ and fix $\sigma_1 \in ]0,\sigma]$ so that

$$\sigma_1\|l\|_{1,\rho,\sigma} \le \frac{1}{4}(\rho - r_1) = \frac{1}{8}(\rho - \hat r). \tag{4.8}$$

We are going to prove that $\Phi(\cdot,t,\varepsilon)$ is a diffeomorphism from $G_{r_1}$ onto its image, if $t \in [0,1]$ and $\varepsilon \in [0,\sigma_1]$. For $\varepsilon = 0$ we have $\Phi(x,t,\varepsilon) = x$, so we can assume that $\varepsilon > 0$. Using the Cauchy integral formula, one gets

$$\varepsilon\Big\|\frac{\partial l}{\partial x}(\cdot,\varepsilon)\Big\|_{r_1} \le \frac{\sigma_1}{\rho - r_1}\,\|l(\cdot,\varepsilon)\|_\rho \le \frac{\sigma_1}{\rho - r_1}\,\|l\|_{1,\rho,\sigma} \le \frac{1}{4}.$$

Hence the matrix

$$M = \frac{\partial \Phi}{\partial x}(x,t,\varepsilon) = I + \varepsilon\chi(t)\,\frac{\partial l}{\partial x}(x,\varepsilon\chi(t)) \tag{4.9}$$

satisfies $|M - I| \leq \frac{1}{4}$. As a consequence, $M$ has an inverse and therefore $\Phi(\cdot, t, \varepsilon)$ is a local diffeomorphism from $G_{r_1}$ onto its image, which is contained in $G_{\frac{3}{4}r_1+\frac{1}{4}\rho}$, the latter by (4.8). If $x_1, x_2 \in G_{r_1}$, then

$$
\begin{aligned}
|\Phi(x_1,t,\varepsilon) - \Phi(x_2,t,\varepsilon)| &= \Big| x_1 - x_2 + \varepsilon\chi(t) \Big( \int_0^1 \frac{\partial l}{\partial x}(\lambda x_1 + (1-\lambda)x_2, \varepsilon\chi(t))\, d\lambda \Big)(x_1 - x_2) \Big| \\
&\geq |x_1 - x_2| - \frac{1}{2}|x_1 - x_2| \\
&= \frac{1}{2}|x_1 - x_2|;
\end{aligned}
$$

note that here the convexity of $G$ (and hence $G_{r_1}$) has been used. It follows that $\Phi(\cdot, t, \varepsilon)$ is one-to-one on $G_{r_1}$ and its inverse $\Psi(\cdot, t, \varepsilon)$ has Lipschitz constant 2. Observe that (4.8) also implies that

$$\sigma_1 \|l\|_{1,\rho,\sigma} \leq \frac{1}{4}(\rho - \hat{r}) = \frac{1}{2}(r_1 - \hat{r}).$$

Arguing analogously to Corollary 3.7, it follows that

$$\Phi(G_{r_1}, t, \varepsilon) \supset G_{\hat{r}}$$

for $t \in [0,1]$ and $\varepsilon \in [0, \sigma_1]$. The Hamiltonian function $h_{\text{aux}}$ from Lemma 4.1 will be defined on the domain

$$\mathcal{D} = \{(X,t,\varepsilon) : t \in [0,1], X \in \Phi(G_{r_1}, t, \varepsilon), \varepsilon \in [0, \sigma_1]\} \supset G_{\hat{r}} \times [0,1] \times [0,\sigma_1]. \tag{4.10}$$

Next we choose the number $\hat{\sigma} \in ]0, \sigma_1]$ so that

$$\hat{\sigma}\|l\|_{1,\rho,\sigma} < \hat{r} - r,$$

which in turn implies that

$$\Phi(G_r, t, \varepsilon) \subset G_{\hat{r}} \tag{4.11}$$

for $t \in [0,1]$ and $\varepsilon \in [0, \hat{\sigma}]$, and moreover we have $\Phi(x,t,\varepsilon) = P_{\varepsilon\chi(t)}(x)$ by definition.

From now on we consider $\Phi$ on $G_{r_1} \times [0,1] \times [0,\hat{\sigma}]$ and the inverse $\Psi(\cdot,t,\varepsilon) = \Phi(\cdot,t,\varepsilon)^{-1}$ has domain $\Phi(G_{r_1}, t, \varepsilon)$. Since $\Phi$ is continuous in its three arguments, the same can be said about $\Psi$. In addition, by the inverse function theorem, $\Psi$ is holomorphic in the first variable. Let $\varepsilon \in [0, \hat{\sigma}]$ be fixed. We will prove that $\Phi(\cdot,\cdot,\varepsilon)$ is $C^1$ in $G_{r_1} \times [0,1]$. Moreover, the cross derivatives do exist, they are continuous and coincide. To see this, we can once again restrict our attention to $\varepsilon > 0$. Since $l \in \mathcal{M}_{1,\rho,\sigma}$, the functions $l(\cdot,\varepsilon)$ and $\frac{\partial l}{\partial \varepsilon}(\cdot,\varepsilon)$ are holomorphic. Hence, by Cauchy's integral formula,

$$\Big\| \frac{\partial l}{\partial x}(\cdot,\varepsilon) \Big\|_{r_1} \leq \frac{1}{\rho - r_1}\|l\|_{1,\rho,\sigma}, \tag{4.12}$$

$$\Big\| \frac{\partial^2 l}{\partial x \partial \varepsilon}(\cdot,\varepsilon) \Big\|_{r_1} \leq \frac{1}{\rho - r_1}\|l\|_{1,\rho,\sigma}. \tag{4.13}$$

Note that in (4.12) the case $\varepsilon = 0$ is admissible. By definition, $\Phi(\cdot,\cdot,\varepsilon)$ is $C^\infty$ in $G_\rho \times ]0,1]$. For $t = 0$, $\Phi(x,0,\varepsilon) = x$ and $\frac{\partial \Phi}{\partial x}(x,0,\varepsilon) = I$. From (4.12) and (4.9) we conclude that $\frac{\partial \Phi}{\partial x}(\cdot,\cdot,\varepsilon)$ is continuous in $G_{r_1} \times [0,1]$. To analyze the derivative w.r. to $t$, we observe that

$$\frac{\partial \Phi}{\partial t}(x,0,\varepsilon) = \lim_{t\to 0^+} \frac{\Phi(x,t,\varepsilon) - \Phi(x,0,\varepsilon)}{t} = \varepsilon \lim_{t\to 0^+} \frac{\chi(t)}{t}\, l(x,\varepsilon\chi(t)) = 0,$$

where we used that $\chi(0) = \dot\chi(0) = 0$ and $\|l\|_{\rho,\sigma} < \infty$. For $t > 0$,

$$\frac{\partial \Phi}{\partial t}(x,t,\varepsilon) = \varepsilon\dot\chi(t)\left[l(x,\varepsilon\chi(t)) + \varepsilon\chi(t)\,\frac{\partial l}{\partial \varepsilon}(x,\varepsilon\chi(t))\right]. \tag{4.14}$$

Thus the continuity of $\frac{\partial \Phi}{\partial t}(\cdot,\cdot,\varepsilon)$ is a consequence of $\|l\|_{1,\rho,\sigma} < \infty$. To summarize, so far we have shown that $\Phi(\cdot,\cdot,\varepsilon)$ is $C^1$ in $G_{r_1} \times [0,1]$. For the cross derivatives, from $\frac{\partial \Phi}{\partial t}(x,0,\varepsilon) = 0$ we deduce that $\frac{\partial^2 \Phi}{\partial x \partial t}(x,0,\varepsilon) = 0$. Also, using (4.9) and (4.12),

$$\frac{\partial^2 \Phi}{\partial t \partial x}(x,0,\varepsilon) = \lim_{t\to 0^+} \frac{\frac{\partial \Phi}{\partial x}(x,t,\varepsilon) - \frac{\partial \Phi}{\partial x}(x,0,\varepsilon)}{t} = \varepsilon \lim_{t\to 0^+} \frac{\chi(t)}{t}\,\frac{\partial l}{\partial x}(x,\varepsilon\chi(t)) = 0.$$

Hence the cross derivatives exist at $t = 0$ and they coincide. The continuity of these derivatives is obtained after differentiating (4.9) w.r. to $t$ in $G_{r_1} \times ]0,1]$; again the bounds (4.12) and (4.13) need to be used here. Both functions $l$ and $h$ belong to the class $\mathcal{M}_{1,\rho,\sigma}$. Thus the previous discussions also apply to the function $\eta(\cdot,\cdot,\varepsilon)$.

Altogether, we see that the relaxed version of Lemma 4.1 can be used to deduce the existence of a function $h_{\rm aux} = h_{\rm aux}(X,t,\varepsilon)$, which is defined on $\mathcal{D}$ from (4.10), with the stated properties. In particular, $h_{\rm aux}(\cdot,t,\varepsilon)$ is well-defined on $G_{\hat r}$. Moreover, if $x \in G_r$, then $X(t) = \Phi(x,t,\varepsilon)$ solves

$$\dot X(t) = J\nabla h_{\rm aux}(X(t),t,\varepsilon) \tag{4.15}$$

by Remark 4.2(a), and also $\varepsilon \in [0,\hat\sigma]$ yields $X(t) \in G_{\hat r}$ for $t \in [0,1]$ due to (4.11). The Poincaré map of (4.15) is $G_r \ni x \mapsto \Phi(x,1,\varepsilon) = x + \varepsilon l(x,\varepsilon) = P_\varepsilon(x)$, i.e., the original map restricted to $G_r$.

To express $h_{\rm aux}$ more explicitly, we recall from the previous computations that

$$\frac{\partial \Phi}{\partial t}(x,t,\varepsilon) = \begin{cases} \varepsilon\dot\chi(t)\,[l(x,\varepsilon\chi(t)) + \varepsilon\chi(t)\,\frac{\partial l}{\partial \varepsilon}(x,\varepsilon\chi(t))] & : \quad t \in ]0,1], \varepsilon \in ]0,\hat\sigma] \\ 0 & : \quad t = 0 \text{ or } \varepsilon = 0 \end{cases} \tag{4.16}$$

and similarly

$$\frac{\partial \eta}{\partial t}(x,t,\varepsilon) = \begin{cases} \varepsilon\dot\chi(t)\,\frac{\partial h}{\partial \varepsilon}(x,\varepsilon\chi(t)) & : \quad t \in ]0,1], \varepsilon \in ]0,\hat\sigma] \\ 0 & : \quad t = 0 \text{ or } \varepsilon = 0 \end{cases}. \tag{4.17}$$

In the notation of Lemma 4.1 we have

$$\mathcal{F}(x,t,\varepsilon) = q + \varepsilon\chi(t)\, l_1(x,\varepsilon\chi(t)), \quad \mathcal{G}(x,t,\varepsilon) = p + \varepsilon\chi(t)\, l_2(x,\varepsilon\chi(t)),$$

where $x = (q,p)$, and $l = (l_1,l_2)$ are the components. Also observe that by (4.3)

$$h_{\rm aux}(X,t,\varepsilon) = \frac{\partial \mathcal{F}}{\partial t}(\Psi(X,t,\varepsilon),t,\varepsilon)\,\mathcal{G}(\Psi(X,t,\varepsilon),t,\varepsilon) - \frac{\partial \eta}{\partial t}(\Psi(X,t,\varepsilon),t,\varepsilon).$$

From $\chi(0)=\dot\chi(0)=\dot\chi(1)=0$ and (4.17) it follows that $h_{\rm aux}(X,t,\varepsilon)=0$ for $t=0$ or $t=1$ or $\varepsilon=0$. Moreover, if $t\neq 0$ and $\varepsilon>0$,

$$\begin{aligned}\frac{1}{\varepsilon}\,h_{\rm aux}(X,t,\varepsilon) &= \dot\chi(t)\Big(l_1(x,\varepsilon\chi(t))+\varepsilon\chi(t)\,\frac{\partial l_1}{\partial\varepsilon}(x,\varepsilon\chi(t))\Big)\Big(p+\varepsilon\chi(t)\,l_2(x,\varepsilon\chi(t))\Big)\\ &\quad -\dot\chi(t)\,\frac{\partial h}{\partial\varepsilon}(x,\varepsilon\chi(t)),\end{aligned}\tag{4.18}$$

and we write $X=(Q,P)$ as well as $x=\Psi(X,t,\varepsilon)$. To pass to the limit $\varepsilon\to 0$ in (4.18), we first recall that $\Psi$ is continuous on $G_{\hat r}\times[0,1]\times[0,\hat\sigma]$ and $\Psi(X,t,0)=X$. From (2.6) in Definition 2.3 of an E-symplectic family we know that $\frac{\partial h}{\partial\varepsilon}(x,\varepsilon)\to\mathfrak{m}(x)$ as $\varepsilon\to 0$ uniformly in $x\in G_\rho$. Thus (4.18) yields

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\,h_{\rm aux}(X,t,\varepsilon)=\dot\chi(t)\,[l_1(X,0)P-\mathfrak{m}(X)]$$

and this limit is uniform in $X\in G_{\hat r}$, $t\in[0,1]$.

Now we define

$$H_{\rm aux}(X,t,\varepsilon)=\left\{\begin{array}{ccc}\frac{1}{\varepsilon}\,h_{\rm aux}(X,t,\varepsilon) & : & \varepsilon\in]0,\hat\sigma]\\ \dot\chi(t)\,[l_1(X,0)P-\mathfrak{m}(X)] & : & \varepsilon=0\end{array}\right. .\tag{4.19}$$

for $X\in G_{\hat r}$ and $t\in[0,1]$, and we are going to verify that $H_{\rm aux}$ has the desired properties. From the above discussions we know that $H_{\rm aux}$ is continuous and

$$H_{\rm aux}(X,0,\varepsilon)=H_{\rm aux}(X,1,\varepsilon)=0.\tag{4.20}$$

As a consequence, $H_{\rm aux}$ can be extended to $G_{\hat r}\times\mathbb{R}\times[0,\hat\sigma]$ in a $T=1$ periodic fashion. First we need to prove that $H_{\rm aux}\in\mathcal{H}_{\hat r,\hat\sigma}$, cf. Definition 3.1. Here (a)-(c) in this definition are straightforward to check. Concerning (d), for $\varepsilon>0$ we know from (4.2) that

$$J\nabla H_{\rm aux}(X,t,\varepsilon)=\frac{1}{\varepsilon}\frac{\partial\Phi}{\partial t}(\Psi(X,t,\varepsilon),t,\varepsilon).$$

Thus, by (4.16),

$$\lim_{\varepsilon\to 0}J\nabla H_{\rm aux}(X,t,\varepsilon)=\dot\chi(t)\,l(X,0),$$

and this limit is uniform in $G_{\hat r}\times\mathbb{R}$. On the other hand, the definition of $H_{\rm aux}$ and (2.8) implies that

$$\begin{aligned}J\nabla H_{\rm aux}(X,t,0) &= \dot\chi(t)\,J\left[P\,\nabla l_1(X,0)-\nabla m(X)+\begin{pmatrix}0\\ l_1(X,0)\end{pmatrix}\right]\\ &= \dot\chi(t)\,J\begin{pmatrix}-l_2(X,0)\\ l_1(X,0)\end{pmatrix}=\dot\chi(t)\,l(X,0).\end{aligned}$$

This shows that $\nabla_X H_{\rm aux}$ is continuous in all of its arguments. Then the bound on $\|\nabla_X H_{\rm aux}\|_{\hat r,\hat\sigma}$ is not difficult to derive from (4.14).

Lastly, we have to establish (4.6). In view of the definition of $H_{\rm aux}$ and (4.20), it suffices to consider $X\in G_{\hat r}$, $t\in]0,1]$ and $\varepsilon\in]0,\hat\sigma]$. From (4.18) we deduce

$$|H_{\rm aux}(X,t,\varepsilon)-H_{\rm aux}(X,t,0)|\le\|\dot\chi\|_\infty\,(R_1+R_2+R_3),$$

where

$$\begin{aligned} R_1 &= |l_1(x,\varepsilon\chi(t))p - l_1(X,0)P|, \\ R_2 &= \Big|\mathfrak{m}(X) - \frac{\partial h}{\partial \varepsilon}(x,\varepsilon\chi(t))\Big|, \\ R_3 &= \varepsilon\chi(t)\,|l_1(x,\varepsilon\chi(t))|\,|l_2(x,\varepsilon\chi(t))| + \varepsilon\chi(t)\left|\frac{\partial l_1}{\partial\varepsilon}(x,\varepsilon\chi(t))\right||p| \\ &\quad + \varepsilon^2\chi(t)^2\left|\frac{\partial l_1}{\partial\varepsilon}(x,\varepsilon\chi(t))\right||l_2(x,\varepsilon\chi(t))|. \end{aligned}$$

For $R_1$, we observe that by definition of $X = \Phi(x,t,\varepsilon)$, see (4.7),

$$|X - x| = \varepsilon\,\chi(t)\,|l(x,\varepsilon\chi(t))| \le \varepsilon\,\|l\|_{1,\rho,\sigma}.$$

Also note that $x = \Psi(X,t,\varepsilon) \in G_{r_1}$ by construction. Therefore

$$|l_1(x,\varepsilon\chi(t)) - l_1(X,\varepsilon\chi(t))| \le 2\left\|\frac{\partial l_1}{\partial x}\right\|_{r_1,\hat\sigma}|x - X| \le \frac{2\varepsilon}{\rho - r_1}\|l_1\|_{r_1,\hat\sigma}\|l\|_{1,\rho,\sigma} \le \frac{2\varepsilon}{\rho - r_1}\|l\|^2_{1,\rho,\sigma}. \tag{4.21}$$

Since $l \in \mathcal{M}_{1,\rho,\sigma}$, also

$$|l(X,\varepsilon\chi(t)) - l(X,0)| \le \|l\|_{1,\rho,\sigma}\,\varepsilon$$

is verified. At this point we need to invoke the geometry of $G = \mathbb{R}\times I$. If $I$ is contained in $[-R,R]$, then $|P| \le R + \hat r \le R + \rho$ as well as $|p| \le R + r_1 \le R + \rho$, due to $X \in G_{\hat r}$ and $x \in G_{r_1}$. Thus altogether, using the foregoing estimates,

$$\begin{aligned} |R_1| &\le |l_1(x,\varepsilon\chi(t))|\,|p - P| + |l_1(x,\varepsilon\chi(t)) - l_1(X,\varepsilon\chi(t))|\,|P| + |l_1(X,\varepsilon\chi(t)) - l_1(X,0)|\,|P| \\ &\le \varepsilon\,\|l\|^2_{1,\rho,\sigma} + \frac{2(R+\rho)\varepsilon}{\rho - r_1}\|l\|^2_{1,\rho,\sigma} + (R+\rho)\,\|l\|_{1,\rho,\sigma}\,\varepsilon, \end{aligned}$$

which is acceptable. For $R_2$ we can argue as follows. Since also $h \in \mathcal{M}_{1,\rho,\sigma}$, we obtain as in (4.21) that

$$\left|\frac{\partial h}{\partial\varepsilon}(x,\varepsilon\chi(t)) - \frac{\partial h}{\partial\varepsilon}(X,\varepsilon\chi(t))\right| \le \frac{2\varepsilon}{\rho - r_1}\left\|\frac{\partial h}{\partial\varepsilon}(\cdot,\varepsilon\chi(t))\right\|_{r_1}\|l\|_{1,\rho,\sigma} \le \frac{2\varepsilon}{\rho - r_1}\|h\|_{1,\rho,\sigma}\|l\|_{1,\rho,\sigma};$$

observe that $\varepsilon\chi(t) \in ]0,\sigma]$ for $\varepsilon \in ]0,\hat\sigma]$ and $t \in ]0,1]$. If we combine this estimate with (2.6), then $R_2 \le C\varepsilon$ is found. Finally, from $l \in \mathcal{M}_{1,\rho,\sigma}$ and $|p| \le R + \rho$, also $R_3 \le C\varepsilon$ is obtained. This completes the argument for (4.6), and hence the proof of the lemma. □

Now we are in a position to complete the

**Proof of Theorem 1.1:** Let $r_2 = \frac{\rho}{3}$ and $r_1 = \frac{2\rho}{3}$. Then Lemma 4.3 can be applied to $l$ and $0 < r_2 < r_1 < \rho$. We deduce that there exist $\sigma_1 \in ]0,\sigma[$ and a Hamiltonian $H_{\rm aux} \in \mathcal{H}_{r_1,\sigma_1}$ such that for $\varepsilon \in [0,\sigma_1]$ the Poincaré map of

$$\dot x = \varepsilon J\nabla H_{\rm aux}(x,t,\varepsilon) \tag{4.22}$$

is $P_\varepsilon$, restricted to $G_{r_2}$. In addition, one can find a constant $C_{\rm aux} > 0$ so that

$$|H_{\rm aux}(x,t,\varepsilon) - H_{\rm aux}(x,t,0)| \le C_{\rm aux}\,\varepsilon \tag{4.23}$$

for $x \in G_{r_1}$, $t \in [0,1]$ and $\varepsilon \in [0,\sigma_1]$. The constant $C_{\rm aux}$ depends upon $\rho$, $\sigma$, $\|l\|_{1,\rho,\sigma}$, the interval $I$, $\|h\|_{1,\rho,\sigma}$ and $\sup_{\varepsilon\in]0,\sigma]} \|\varepsilon^{-1}(\frac{\partial h}{\partial\varepsilon}(\cdot,\varepsilon) - \mathfrak{m})\|_\rho$.

Next we are going to invoke Theorem 3.14 to $H_{\rm aux}$ for the parameters $r = r_2$, $\rho = r_1$, $\sigma = \sigma_1$ and $T = 1$. By this result, we can find $C, D > 0$ (depending upon $\rho$ and $\|\nabla H_{\rm aux}\|_{r_1,\sigma_1}$) with the following properties. There is a change of variables $x = \Gamma(y,t,\varepsilon)$, which is defined on $G_{r_2} \times \mathbb{R} \times [0,\sigma_1]$ and which satisfies $\Gamma(G_{r_2},t,\varepsilon) \subset G_{r_1}$ for $t \in \mathbb{R}$ and $\varepsilon \in [0,\sigma_1]$, such that (4.22) is transformed into

$$\dot y = \varepsilon(J\nabla_y \mathcal{N}(y,\varepsilon) + J\nabla_y \mathcal{R}(y,t,\varepsilon)), \tag{4.24}$$

for functions $\mathcal{N} \in \mathcal{H}_{r_2,\sigma_1}$ and $\mathcal{R} \in \widetilde{\mathcal{H}}_{r_2,\sigma_1}$. Furthermore,

$$\|\nabla_y \mathcal{N}(\cdot,\varepsilon)\|_{r_2} \le C, \tag{4.25}$$

$$\|\nabla_y \mathcal{R}(\cdot,\cdot,\varepsilon)\|_{r_2} \le Ce^{-D/\varepsilon}, \tag{4.26}$$

$$|\mathcal{N}(y,\varepsilon) - \mathcal{N}(y,0)| \le C\varepsilon + \|\overline{H_{\rm aux}}(\cdot,\varepsilon) - \overline{H_{\rm aux}}(\cdot,0)\|_{r_1}, \tag{4.27}$$

for $y \in G_{r_2}$ and $\varepsilon \in [0,\sigma_1]$. In addition,

$$\mathcal{N}(y,0) = \overline{H_{\rm aux}}(y,0) \tag{4.28}$$

is verified. According to the definition of $H_{\rm aux}$ in (4.19) and by (4.18), one sees that it is possible to bound $\|\nabla H_{\rm aux}\|_{r_1,\sigma_1}$ in terms of $\|l\|_{1,\rho,\sigma}$, the interval $I$ and $\|h\|_{1,\rho,\sigma}$.

For later reference we first discuss the connection between $\mathcal{N}$ and the function $E$ from Theorem 1.1, cf. (2.10), and we also consider the variation of $\mathcal{N}$ w.r. to $\varepsilon$. From (4.28), the definition of $H_{\rm aux}(y,0)$ in (4.19) and (2.10),

$$\mathcal{N}(y,0) = \overline{H_{\rm aux}}(y,0) = \int_0^1 \dot\chi(t)\,[l_1(y,0)P - \mathfrak{m}(y)]\,dt = l_1(y,0)P - \mathfrak{m}(y) = E(y), \tag{4.29}$$

where $y = (Q,P)$. Using (4.27) and (4.23), we moreover find for $y \in G_{r_2}$ and $\varepsilon \in [0,\sigma_1]$ that

$$|\mathcal{N}(y,\varepsilon) - \mathcal{N}(y,0)| \le C\varepsilon + \|\overline{H_{\rm aux}}(\cdot,\varepsilon) - \overline{H_{\rm aux}}(\cdot,0)\|_{r_1} \le C_1\varepsilon, \tag{4.30}$$

where the constant $C_1 = C + C_{\rm aux}$ depends upon $\rho$, $\sigma$, $\|l\|_{1,\rho,\sigma}$, the interval $I$, $\|h\|_{1,\rho,\sigma}$ and $\sup_{\varepsilon\in]0,\sigma]} \|\varepsilon^{-1}(\frac{\partial h}{\partial\varepsilon}(\cdot,\varepsilon) - \mathfrak{m})\|_\rho$; henceforth all constants are allowed to depend upon those parameters.

Now we define $r_3 = \frac{2r_2}{3} = \frac{2\rho}{9}$ and $r_4 = \frac{r_2}{3} = \frac{\rho}{9}$ to obtain $0 < r_4 < r_3 < r_2 < r_1 < \rho$. According to Corollary 3.15 there is $\sigma_2 \in ]0,\sigma_1]$ such that

$$G_{r_3} \subset \Gamma(G_{r_2},t,\varepsilon)$$

for $t \in \mathbb{R}$ and $\varepsilon \in [0,\sigma_2]$; in particular, $\Gamma(\cdot,t,\varepsilon)^{-1} : G_{r_3} \to G_{r_2}$ is well-defined.

Let $\Phi(x,t,\varepsilon)$ denote the solution to (4.22) satisfying $\Phi(x,0,\varepsilon) = x$. Similarly, $\phi(y,t,\varepsilon)$ will be used for the solution to (4.24) so that $\phi(y,0,\varepsilon) = y$. Now we select $\sigma_3 \in ]0,\sigma_2]$ such that

$\Phi(x,t,\varepsilon)$ is well-defined on $G_{r_4}\times[0,1]\times[0,\sigma_2]$ and takes values in $G_{r_3}$. The solutions of the two systems are connected by the formula

$$\phi(y,t,\varepsilon)=\Gamma^{-1}(\Phi(\Gamma(y,0,\varepsilon),t,\varepsilon),t,\varepsilon)=\Gamma^{-1}(\Phi(y,t,\varepsilon),t,\varepsilon)$$

for $y\in G_{r_4}$, $t\in[0,1]$ and $\varepsilon\in[0,\sigma_3]$. Letting $t=1$ and taking into account condition (b) in Definition 3.13, it follows that

$$\phi(y,1,\varepsilon)=\Gamma^{-1}(\Phi(y,1,\varepsilon),1,\varepsilon)=\Phi(y,1,\varepsilon)=P_\varepsilon(y).$$

In other words, $P_\varepsilon$ is also the Poincaré map of (4.24), at least in the domain $G_{r_4}$.

Now we are going to consider the autonomous system

$$\dot y=\varepsilon J\nabla_y\mathcal{N}(y,\varepsilon), \tag{4.31}$$

denoting by $\hat\phi(y,t,\varepsilon)$ the associated flow. Using (4.25), we deduce that there is $\hat\sigma\in]0,\sigma_3]$ with the property that $\hat\phi(y,t,\varepsilon)$ is well-defined on $G_{r_4}\times[0,1]\times[0,\hat\sigma]$ and moreover

$$\hat\phi(G_{r_4}\times[0,1]\times[0,\hat\sigma])\subset G_{r_3}.$$

The system (4.31) is Hamiltonian, with Hamiltonian function $\varepsilon\mathcal{N}(\cdot,\varepsilon)$. In particular, if $\hat P_\varepsilon=\hat\phi(\cdot,y,1)$ denotes the Poincaré map of (4.31), then

$$\mathcal{N}(\hat P_\varepsilon(y),\varepsilon)=\mathcal{N}(y,\varepsilon),\quad y\in G_{r_4},\quad \varepsilon\in[0,\hat\sigma]. \tag{4.32}$$

To estimate the difference between $\phi$ and $\hat\phi$, we first observe that for $\varepsilon\in[0,\hat\sigma]$,

$$\|D^2\mathcal{N}(\cdot,\varepsilon)\|_{r_3}\le\frac{1}{r_3-r_2}\|\nabla\mathcal{N}(\cdot,\varepsilon)\|_{r_2}\le\frac{1}{r_3-r_2}C=C_2,$$

where we have once again resorted to (4.25). If $(y,t,\varepsilon)\in G_{r_4}\times[0,1]\times[0,\hat\sigma]$, then the systems (4.24), (4.31) in conjunction with (4.26) yield

$$\begin{aligned}|\phi(y,t,\varepsilon)-\hat\phi(y,t,\varepsilon)| &= \varepsilon\Big|\int_0^t[J\nabla_y\mathcal{N}(\phi(y,s,\varepsilon),\varepsilon)+J\nabla_y\mathcal{R}(\phi(y,s,\varepsilon),s,\varepsilon)\\ &\qquad -J\nabla_y\mathcal{N}(\hat\phi(y,s,\varepsilon),\varepsilon)]\,ds\Big|\\ &\le C_2\,\varepsilon\int_0^t|\phi(y,s,\varepsilon)-\hat\phi(y,s,\varepsilon)|\,ds+C\varepsilon e^{-D/\varepsilon}.\end{aligned}$$

Hence from Gronwall's inequality,

$$|\phi(y,t,\varepsilon)-\hat\phi(y,t,\varepsilon)|\le C\varepsilon e^{-D/\varepsilon}\,e^{C_2\varepsilon t}.$$

For the Poincaré maps, i.e., at $t=1$, we deduce

$$|P_\varepsilon(y)-\hat P_\varepsilon(y)|\le C_3\,\varepsilon\,e^{-D/\varepsilon},\quad y\in G_{r_4},\quad \varepsilon\in[0,\hat\sigma], \tag{4.33}$$

where $C_3=Ce^{C_2\hat\sigma}$.

Now we are ready to complete the proof. Let $(x_n)_{0\le n\le N} = (P_\varepsilon^n(x_0))_{0\le n\le N}$ be a real forward orbit piece of $P_\varepsilon$ so that $x_n \in G$ for all $0 \le n \le N$. Since $G \subset G_{r_4}$, all the previous properties can be used along the orbit. From (4.29) and (4.30) we get

$$\begin{aligned} |E(x_n) - E(x_0)| &\le |E(x_n) - \mathcal{N}(x_n,\varepsilon)| + |\mathcal{N}(x_n,\varepsilon) - \mathcal{N}(x_0,\varepsilon)| + |\mathcal{N}(x_0,\varepsilon) - E(x_0)| \\ &= |\mathcal{N}(x_n,0) - \mathcal{N}(x_n,\varepsilon)| + |\mathcal{N}(x_n,\varepsilon) - \mathcal{N}(x_0,\varepsilon)| + |\mathcal{N}(x_0,\varepsilon) - \mathcal{N}(x_0,0)| \\ &\le 2C_1\varepsilon + |\mathcal{N}(x_n,\varepsilon) - \mathcal{N}(x_0,\varepsilon)|. \end{aligned}$$

In addition, (4.32), (4.25) and (4.33) lead to

$$\begin{aligned} |\mathcal{N}(x_n,\varepsilon) - \mathcal{N}(x_0,\varepsilon)| &\le \sum_{j=0}^{n-1} |\mathcal{N}(P_\varepsilon(x_j),\varepsilon) - \mathcal{N}(x_0,\varepsilon)| \\ &= \sum_{j=0}^{n-1} |\mathcal{N}(P_\varepsilon(x_j),\varepsilon) - \mathcal{N}(\hat{P}_\varepsilon(x_0),\varepsilon)| \\ &\le CC_3\, n\, \varepsilon\, e^{-D/\varepsilon}. \end{aligned}$$

Thus the claim follows if we define $\hat{C} = 2C_1 + CC_3$ and $\hat{D} = D$. □

# References


[1] Arnold V.I.: *Mathematical Methods of Classical Mechanics*, 2nd edition, Springer, Berlin-New York 1989

[2] Benettin G. & Giorgilli A.: On the Hamiltonian interpolation of near-to-the-identity symplectic mappings with application to symplectic integration algorithms, *J. Statist. Phys.* **74**, 1117-1143 (1994)

[3] Kunze M. & Ortega R.: Non-periodic twist maps, in *Stability and Bifurcation Theory for Non-Autonomous Differential Equations*, Eds. Johnson R. & Pera M.P., Lecture Notes in Mathematics Vol. 2065, Springer, Berlin-New York 2013, pp. 265-300

[4] Kunze M. & Ortega R.: Growth rates of orbits in non-periodic twist maps and a theorem by Neishtadt, preprint 2020

[5] Meyer K. & Offin D.: *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*, 3rd edition, Springer, Berlin-New York 2017

[6] Neishtadt A.I.: The separation of motions in systems with rapidly rotating phase, *J. Appl. Math. Mech.* **48**, 133-139 (1985); translated from *Prikl. Mat. Mekh.* **48**, 197-204 (1984)

[7] Poincaré H.: *Leçons de Mécanique Céleste*, t. I, Gauthier-Villars, Paris 1905

[8] Treschev D. & Zubelevich O.: *Introduction to the Perturbation Theory of Hamiltonian Systems*, Springer, Berlin-New York 2010