This is an author manuscript post-peer-reviewing (accepted version) of the original publication. The layout of the published version may differ .

––

Explicit, time reversible, adaptive step size control

––

Hairer, Ernst; Soederlind, Gustaf

This publication URL: https://archive-ouverte.unige.ch/unige:12120

EXPLICIT, TIME REVERSIBLE, ADAPTIVE STEP SIZE CONTROL[*]

ERNST HAIRER[†] AND GUSTAF SÖDERLIND[‡]

Abstract. Adaptive step size control is difficult to combine with geometric numerical integration. As classical step size control is based on "past" information only, time symmetry is destroyed and with it the qualitative properties of the method. In this paper we develop completely explicit, reversible, symmetry-preserving, adaptive step size selection algorithms for geometric numerical integrators such as the Störmer–Verlet method. A new step density controller is proposed and analyzed using backward error analysis and reversible perturbation theory. For integrable reversible systems we show that the resulting adaptive method nearly preserves all action variables and, in particular, the total energy for Hamiltonian systems. It has the same excellent long-term behavior as that obtained when constant steps are used. With variable steps, however, both accuracy and efficiency are greatly improved.

Key words. Adaptive integration, geometric integration, time reversible and symmetric methods, Störmer–Verlet method, Hamiltonian systems, explicit and reversible step size control, backward error analysis, reversible perturbation theory

AMS subject classifications. 65L05, 65P10

DOI. 10.1137/040606995

1. Introduction. Geometric integrators (symplectic methods for Hamiltonian systems, symmetric methods for reversible problems, volume-conserving methods for divergence-free systems, etc.) are known for their excellent behavior when constant step size integration over long times is considered. As first observed by Gladman, Duncan, and Candy [3] and Calvo and Sanz-Serna [1], classical step size strategies destroy these properties. Thus, if step size selection is based on *past* information, symmetry breaks down, because what is "past" depends on the direction of integration. No advantage over explicit Runge-Kutta or multistep methods is then left [4, Chap. VIII].

The remaining possibility is to control step size using *present* information only. Several such attempts have been made. For reversible differential equations, so-called reversible step size strategies were proposed by Hut, Makino, and McMillan [8] and by Stoffer [11]. The step size is defined by an implicit algebraic relation, producing the same result when integrating forward or backward in time.

An explicit step size scheme for the Störmer–Verlet method was proposed in Huang and Leimkuhler [7] and further improved in Holder, Leimkuhler, and Reich [6]. Based on a two-term step size recursion, it is prone to creating undesirable oscillations in step size and numerical solution; see Cirilli, Hairer, and Leimkuhler [2].

In this paper we develop a theory for the construction of completely explicit, symmetric, and time reversible step size selection schemes, which are stable and non-oscillatory. In particular, we propose an integrating controller for the *step density*, see

Söderlind [9, 10] for control theoretic notions. Its excellent long-term performance is illustrated in connection with the Störmer–Verlet method.

The resulting adaptive geometric numerical integrator is analyzed using backward error analysis combined with reversible perturbation theory. For integrable, reversible Hamiltonian systems we prove that there is near-preservation of the Hamiltonian and of all action variables over long times, and that the global error grows only linearly. This equals the best possible qualitative properties that can be obtained with constant step size. Due to the varying steps, however, accuracy and efficiency are significantly improved.

2. Reversible adaptive integration. Let us consider first-order systems of differential equations of the form

$$(2.1) \qquad \begin{aligned} \dot{p} &= f(p, q), \\ \dot{q} &= g(p, q), \end{aligned}$$

with initial conditions at $t = 0$. This system is assumed to be time reversible with respect to the linear involution $S : (p, q) \mapsto (-p, q)$; thus we assume that the functions f and g satisfy the reversibility conditions

$$(2.2) \qquad \begin{aligned} f(-p, q) &= f(p, q), \\ g(-p, q) &= -g(p, q). \end{aligned}$$

An example is Hamiltonian systems in Newtonian dynamics, where q and p represent position and momentum, and $g(p, q) = M^{-1}p$ and $f(p, q) = -\nabla U(q)$ for a mass matrix M and a potential $U(q)$. The approach below, however, applies to more general systems, and p and q need not have the same dimension.

Collecting p and q in a vector $y = (p, q)$ and introducing $F = (f, g)$, the system and the reversibility condition can be written

$$(2.3) \qquad \dot{y} = F(y), \qquad \text{with} \qquad -F = SFS.$$

Assuming that F is Lipschitz, the flow φ_t of (2.3) has inverses $\varphi_t^{-1} = \varphi_{-t}$. Under the reversibility condition it also satisfies $\varphi_t^{-1} = S\varphi_t S$.

2.1. Symmetry and reversibility. A one-step method; $\Phi_h : y_n \mapsto y_{n+1}$ is called *symmetric* if $\Phi_h^{-1} = \Phi_{-h}$ and *reversible* if $\Phi_h^{-1} = S\Phi_h S$, see Figure 2.1. Since
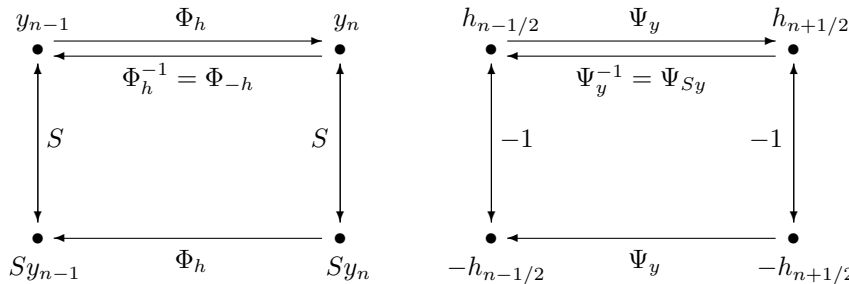

FIG. 2.1. *Commutative diagrams. Left: Upper branch represents the symmetry condition on the numerical method Φ_h, and the lower branch represents reversibility. Right: Upper branch shows reversibility for the step size map Ψ_y, while the lower branch represents symmetry.*

all reasonable methods satisfy $S\Phi_h S = \Phi_{-h}$ (e.g., partitioned Runge–Kutta methods), reversibility is equivalent to symmetry.

To make the method adaptive, we need notions of symmetry and reversibility for the step size control. This requires an invertible step size map $\Psi_{y_n} : h_{n-1/2} \mapsto h_{n+1/2}$, depending on y_n only and where the step size is defined by $h_{n+1/2} = t_{n+1} - t_n$.

DEFINITION 2.1. *An invertible step size map* $\Psi_y : \mathbb{R} \to \mathbb{R}$ *is called* symmetric *if* $\Psi_y^{-1} = -\mathrm{id} \circ \Psi_y \circ (-\mathrm{id})$. *It is called* reversible *with respect to* S *if* $\Psi_y^{-1} = \Psi_{Sy}$.

A symmetric and reversible control map therefore satisfies

$$(2.4) \qquad\qquad \Psi_{Sy} = -\mathrm{id} \circ \Psi_y \circ (-\mathrm{id}).$$

The symmetry condition requires that $-\Psi_y$ is an involution (see Figure 2.1); a symmetric Ψ_y maps $h_{n-1/2}$ to $h_{n+1/2}$, and $-h_{n+1/2}$ to $-h_{n-1/2}$. Further, Ψ_y is nonlinear; otherwise the symmetry condition requires Ψ_y itself to be an involution, which leads to constant steps or oscillatory control. As a consequence, conventional multiplicative step size control of the type $\Psi_y : h_{n-1/2} \mapsto \theta_n \cdot h_{n-1/2}$ is ruled out. Finally, if Ψ_y is reversible, then Ψ_{Sy} maps $h_{n+1/2}$ to $h_{n-1/2}$.

As the combination of method and step size controller must be able to run forward as well as backward in time, it is essential that the controller retain its structure and stability independently of the direction of integration. Figure 2.2 shows that the control map in forward as well as backward time is Ψ_y. Likewise, Figure 2.3 shows that in reversed time the control map is Ψ_{Sy}.

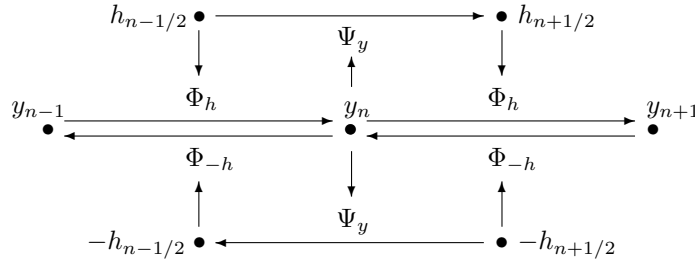FIG. 2.2. *Adaptive, symmetric method integrates forward (upper half) and backwards (lower half) in time. In both cases the same symmetric step size map* Ψ_y *governs* h *and* $-h$, *respectively.*
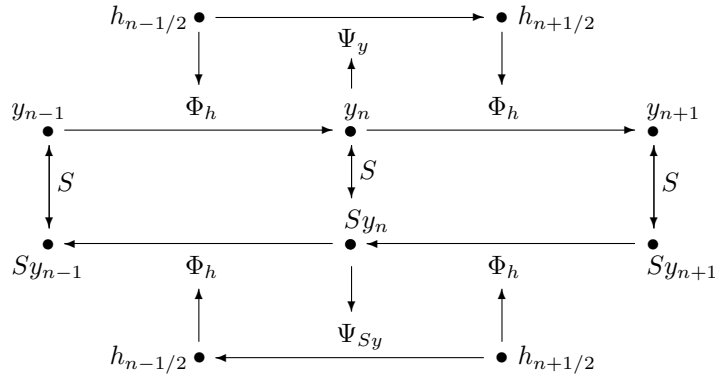

FIG. 2.3. *Adaptive, reversible method integrates in forward (upper half) and reverse (lower half) time. A reversible* Ψ_y *governs* h *in forward time, while* Ψ_{Sy} *controls it in reverse time.*

2.2. Discrete integrating control. In order to construct suitable control maps $\Psi_y : u \mapsto v$, we look for rational functions $R(u, v)$, where numerator and denominator are multinomials, linear in u and v. A unique, explicit solution for either u or v can then be found from the *linear* equation $R(u, v) = G(y)$.

The rational function R represents the *control structure*. To make the controller symmetric, we impose the *involution criterion* $R(u, v) = R(-v, -u)$ on R. The function $G(y)$ is derived from the *control objective*. To make the controller reversible, we impose $R(u, v) = -R(v, u)$ on R, and on G the *reversibility condition*

$$(2.5) \qquad\qquad\qquad\qquad GS = -G.$$

The simplest control structure is an *integrating controller*. This implies that the numerator of R is a difference. There are then only two controllers within the above construction. The elementary choice is $R(u, v) = v - u$, leading to the recursion

$$(2.6) \qquad\qquad\qquad h_{n+1/2} = h_{n-1/2} + G(y_n).$$

This integrating controller is both symmetric and reversible.

The other choice is $R(u, v) = (u - v)/(vu)$, with step size selection schemes

$$(2.7) \qquad \frac{h_{n-1/2} - h_{n+1/2}}{h_{n+1/2} h_{n-1/2}} = G(y_n) \quad \Rightarrow \quad h_{n+1/2} = \frac{h_{n-1/2}}{1 + G(y_n)h_{n-1/2}}.$$

This symmetric and reversible controller will be seen to have several advantages over (2.6). For instance, if $G(y)$ is bounded, the step size changes smoothly: the step size increment is $h_{n+1/2} - h_{n-1/2} = \mathrm{O}(h_{n-1/2}^2)$.

In particular, we propose the following special version of (2.7),

$$(2.8) \qquad\qquad\qquad \rho_{n+1/2} = \rho_{n-1/2} + \varepsilon G(y_n),$$

where ε is constant and the step size is recovered through $h_{n+1/2} = \varepsilon/\rho_{n+1/2}$. This symmetric, reversible, integrating controller will be analyzed in the following sections.

In general, the reversibility condition on G is not sufficient for constructing a good controller. Control structure and objective are strongly linked. A proper G can be found by investigating the continuous analogue of (2.8).

2.3. Continuous integrating control. Introduce a differentiable time transformation $t = \Gamma(\tau)$ with derivative with respect to τ given by

$$(2.9) \qquad\qquad\qquad (\mathrm{D}_\tau \Gamma)(\tau) = \Gamma'(\tau) = 1/\rho(\tau) > 0.$$

Thus prime denotes derivative with respect to τ while dot denotes derivative with respect to t. It follows that $\mathrm{d}t = \mathrm{d}\tau/\rho(\tau)$ and $\mathrm{D}_t = \rho\mathrm{D}_\tau$. Both t and τ will be sampled, with $t_n = \Gamma(\tau_n)$, such that equidistant samples of τ correspond to a nonequidistant grid in t. Thus, defining $\tau_{n+1} - \tau_n = \varepsilon$ for all n, we have

$$(2.10) \qquad h_{n+1/2} = t_{n+1} - t_n = \Gamma(\tau_{n+1}) - \Gamma(\tau_n) \approx \frac{\tau_{n+1} - \tau_n}{\rho(\tau_{n+1/2})} = \frac{\varepsilon}{\rho(\tau_{n+1/2})}.$$

Convergence and error bounds are studied as $\varepsilon \to 0$, while the *step density* $\rho(\tau)$, normalized by the condition $\rho(0) = 1$, accounts for step size variation.

All previous approaches to adaptive geometric integration assume ρ to be a prescribed function of y. In contrast, we shall let ρ be determined dynamically, in actual

computations by the discrete controller (2.8), and here, for analysis purposes, by the corresponding *continuous integrating controller*

$$(2.11) \qquad\qquad \rho' = G(y), \qquad \rho(0) = 1.$$

The original system (2.3) is thus replaced by the augmented system

$$(2.12) \qquad\qquad \begin{aligned} y' &= F(y)/\rho, \\ \rho' &= G(y). \end{aligned}$$

If discretized with constant step ε, the augmented system generates a variable step size discretization for the original system $\dot{y} = F(y)$.

Choice of the function $G(y)$. Conventional adaptivity aims at keeping the product of (some power of) h and a scalar function $Q(y)$ equal to a tolerance. Thus, considering the control objective $Q(y)/\rho = \text{Const}$, we have

$$(2.13) \qquad\qquad \begin{aligned} \rho' &= \nabla Q(y)^{\mathrm{T}} F(y)/Q(y), \\ t' &= 1/\rho, \end{aligned}$$

where the second equation recovers the original time t. Whenever $QS = Q$, the function $G(y) = \nabla Q(y)^{\mathrm{T}} F(y)/Q(y)$ satisfies $GS = -G$. For this particular choice of $G(y)$, (2.13) is *Hamiltonian, with first integral* $Q(y)/\rho$. With the "step size Hamiltonian" $\hbar(\rho, t) = \log[Q(y(t))/\rho]$, the system can be written $\rho' = \hbar_t$; $t' = -\hbar_\rho$.

The proposed *discrete* integrating controller,

$$(2.14) \qquad\qquad \begin{aligned} \rho_{n+1/2} &= \rho_{n-1/2} + \varepsilon\, \nabla Q(y(t_n))^{\mathrm{T}} F(y(t_n))/Q(y(t_n)), \\ t_{n+1} &= t_n + \varepsilon/\rho_{n+1/2}, \end{aligned}$$

can now be viewed as a fixed step size Störmer–Verlet discretization of (2.13): symmetric, reversible, and symplectic. The discrete step density $\{\rho_{n+1/2}\}$ has errors of magnitude ε^2. As the step size Hamiltonian $\hbar(\rho, t)$ is nearly preserved, however, we have $Q(y(t_n))/\rho_n \approx \text{Const}$. The discrete controller is therefore stable.

3. Main results. The established facts are that (i) a geometric integrator has good long-time behavior for constant steps and (ii) the step density controller (2.14) has good long-time behavior for exact data $\{y(t_n)\}$. The problem is now to show that a geometric integrator, generating approximate data $\{y_n\}$ for (2.14) in lieu of $\{y(t_n)\}$, while *governed* by the controller, produces a stable, adaptive integrator with a good long-time behavior. We shall consider the following general algorithm.

ALGORITHM. *Let Φ_h be a one-step method for* (2.3) *with initial value y_0. Further, let $\rho_0 = 1$, and let $\varepsilon > 0$ be constant. Define $\{y_n\}$ by*

$$(3.1) \qquad\qquad \begin{aligned} \rho_{n+1/2} &= \rho_n + \varepsilon\, G(y_n)/2, \\ y_{n+1} &= \Phi_{\varepsilon/\rho_{n+1/2}}(y_n), \\ \rho_{n+1} &= \rho_{n+1/2} + \varepsilon\, G(y_{n+1})/2, \end{aligned}$$

where y_n approximates $y(t_n)$, and $t_{n+1} = t_n + \varepsilon/\rho_{n+1/2}$.

THEOREM 3.1. *For the algorithm* (3.1), *let*

$$(3.2) \qquad\qquad \widehat{\Phi}_\varepsilon : \begin{pmatrix} y_n \\ \rho_n \end{pmatrix} \mapsto \begin{pmatrix} y_{n+1} \\ \rho_{n+1} \end{pmatrix} \qquad and \qquad \widehat{S} = \begin{pmatrix} S & 0 \\ 0 & 1 \end{pmatrix}.$$

It then holds that (i) $\widehat{\Phi}_\varepsilon$ *is symmetric whenever* Φ_h *is symmetric;* (ii) $\widehat{\Phi}_\varepsilon$ *is reversible with respect to* $\widehat{S}$ *whenever* Φ_h *is reversible with respect to* S *and* $GS = -G$.

The algorithm can be interpreted as a Strang splitting for the solution of (2.12): it approximates the flow of (2.11) with fixed y over a half-step $\varepsilon/2$; then applies the method Φ_ε to $y' = F(y)/\rho$ with fixed ρ; finally, it computes a second half-step of (2.11) with fixed y. The recursion generates a sequence $\rho_{n+1/2} \approx \rho(\tau_{n+1/2})$, where the step size is *defined* as $h_{n+1/2} = \varepsilon/\rho_{n+1/2}$. Thus, in spite of not knowing $\rho(\tau)$ exactly, both the time sequence $\{t_n\}$ and the transformed time $\{\tau_n\}$ are explicitly obtained by the time recursions.

3.1. The variable step size Störmer–Verlet method. Consider a Hamiltonian system with separable Hamiltonian $H(p,q) = T(p) + U(q)$. Using the Störmer–Verlet method, the reversible, variable step size algorithm becomes (starting with $\rho_0 = 1$ and $\rho_{1/2} = \rho_0 + \varepsilon\, G(p_0, q_0)/2$ or, equivalently, $\rho_{-1/2} = \rho_0 - \varepsilon\, G(p_0, q_0)/2$)

$$
\begin{aligned}
\rho_{n+1/2} &= \rho_{n-1/2} + \varepsilon\, G(p_n, q_n), \\
p_{n+1/2} &= p_n - \varepsilon\, \nabla U(q_n)/(2\rho_{n+1/2}), \\
q_{n+1} &= q_n + \varepsilon\, \nabla T(p_{n+1/2})/\rho_{n+1/2}, \\
p_{n+1} &= p_{n+1/2} - \varepsilon\, \nabla U(q_{n+1})/(2\rho_{n+1/2}).
\end{aligned}
$$
(3.3)

This method is explicit, symmetric, and reversible as long as $GS = -G$, and it computes approximations on a nonequidistant grid $\{t_n\}$ given by $t_{n+1} = t_n + \varepsilon/\rho_{n+1/2}$.

Let us apply this method to the Kepler problem, which is Hamiltonian with

$$
H(p,q) = \frac{p^{\mathrm{T}} p}{2} - \frac{1}{\sqrt{q^{\mathrm{T}} q}}.
$$
(3.4)

The initial conditions are taken as

$$
q(0) = (1 - e,\, 0)^{\mathrm{T}}; \qquad p(0) = (0,\, \sqrt{(1+e)/(1-e)}\,)^{\mathrm{T}},
$$
(3.5)

where we choose the eccentricity $e = 0.8$. Further, we take $\varepsilon = 0.005$ and select $Q(q) = (q^{\mathrm{T}} q)^{-\alpha/2}$ with $\alpha = 3/2$, so that the control function (see (2.13)) becomes

$$
G(p,q) = -\alpha\, p^{\mathrm{T}} q / q^{\mathrm{T}} q.
$$
(3.6)

Figure 3.1 shows the error in the Hamiltonian along the numerical solution as well as the global error in the solution. The error in the Hamiltonian remains bounded and proportional to ε^2, and the global error grows linearly with time (in double logarithmic scale a linear growth corresponds to a line with slope one; such lines are drawn in grey). This is qualitatively identical to the behavior observed in constant step size implementations of symplectic methods.

Figure 3.2 shows the step sizes $h_{n+1/2} = \varepsilon/\rho_{n+1/2}$ as a function of time. Included is the graph of the control error $Q(p_n, q_n)/\rho_n - Q(p_0, q_0)/\rho_0$ in grey. Since this deviation from the constant value $Q(p_0, q_0)/\rho_0$ is small without any drift, the step density remains close to $Q(p,q)$ multiplied by a constant.

A theoretical explanation of this excellent behavior of the variable step size implementation will be given in the following subsection.
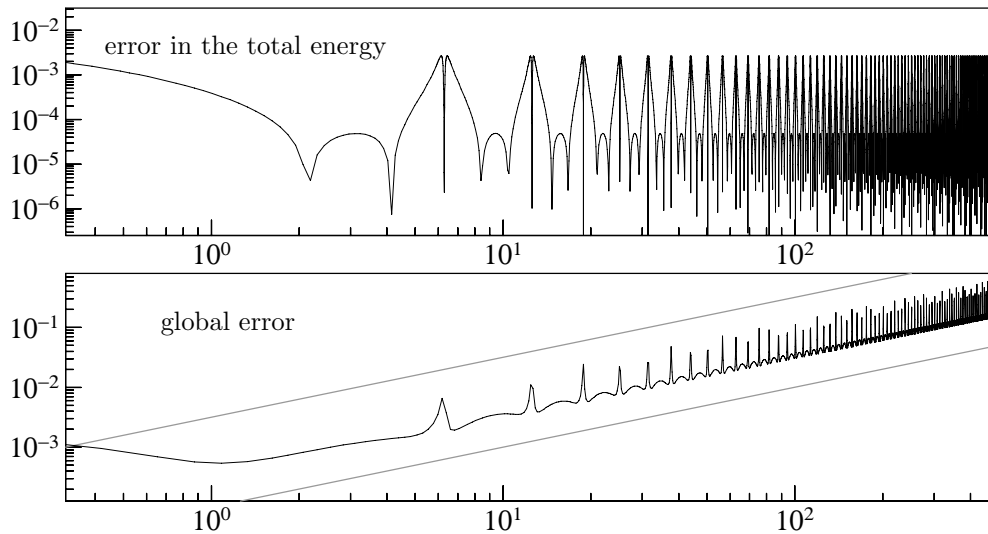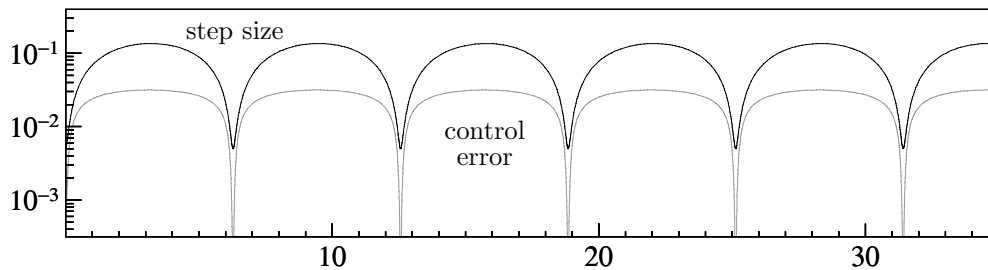
FIG. 3.1. *Numerical Hamiltonian and global error as a function of time.*

FIG. 3.2. *Step sizes of the variable step size Störmer–Verlet method as a function of time, and the control error $Q(p_n, q_n)/\rho_n - Q(p_0, q_0)/\rho_0$ (grey curve).*

3.2. Integrable reversible systems. Consider a differential equation (2.1) satisfying the reversibility conditions (2.2). Such a system is an *integrable reversible system* if there exists a reversibility-preserving transformation

$$(3.7) \qquad (p, q) = \psi(\theta, a), \qquad \psi S = S\psi,$$

to *action-angle variables*, defined for actions $a = (a_1, \ldots, a_m)$ in an open set of $\mathbb{R}^m$ and for angles $\theta = (\theta_1, \ldots, \theta_d)$ on the whole torus

$$\mathbb{T}^d = \mathbb{R}^d/(2\pi\mathbb{Z}^d) = \{(\theta_1, \ldots, \theta_d) : \theta_i \in \mathbb{R} \bmod 2\pi\},$$

such that the transformed system (2.1) is of the form

$$(3.8) \qquad \begin{aligned} \dot{a} &= 0, \\ \dot{\theta} &= \omega(a) \end{aligned}$$

(see [4, Chap. XI] for the connection to completely integrable Hamiltonian systems and examples). Denoting the inverse transformation of (3.7) by

$$(\theta, a) = (\Theta(p, q), I(p, q)),$$

the components of $I = (I_1, \ldots, I_m)$ are even functions of p and first integrals of the system (2.1). For the following result, which shows linear error growth and near-preservation of the action variables over long times for the variable step size algorithm, we need the *Diophantine condition*

$$(3.9) \qquad |k \cdot \omega| \geq \gamma |k|^{-\nu}, \qquad k \in \mathbb{Z}^d, \ k \neq 0,$$

for positive constants γ and ν (cf. [4, sect. X.2.1]). Theorem XI.3.1 of [4] is now extended to the following result for the reversible step size control algorithm.

THEOREM 3.2. *Consider the adaptive method* (3.1) *with symmetric and reversible basic method* Φ_h *of order* $2r$ *and reversible* $G(y) = \nabla Q(y)^{\mathrm{T}} F(y)/Q(y)$, *applied to an integrable reversible system* (2.1) *with real-analytic functions* f, g, G, *and* ψ. *Suppose that* $\omega(a^*)$ *satisfies the Diophantine condition* (3.9). *Then, for an arbitray truncation index* N, *there exist positive constants* C, c, ε_0 *such that the following holds for all* $\varepsilon \leq \varepsilon_0$: *every numerical solution* (p_n, q_n) *starting with* $\|I(p_0, q_0) - a^*\| \leq c\,|\log \varepsilon\|^{-\nu-1}$ *satisfies*

$$\|(p_n, q_n) - (p(t_n), q(t_n))\| \leq C\,t_n\,\varepsilon^{2r} \quad for \quad 0 \leq t_n \leq \varepsilon^{-2r},$$
$$\|I(p_n, q_n) - I(p_0, q_0)\| \leq C\,\varepsilon^{2r} \quad for \quad 0 \leq t_n \leq \varepsilon^{-2(N-r)},$$
$$|Q(p_n, q_n)/\rho_n - Q(p_0, q_0)/\rho_0| \leq C\,\varepsilon^2 \quad for \quad 0 \leq t_n \leq \varepsilon^{-2(N-1)}.$$

The constants ε_0, c, C *depend on* γ, ν, N *and on the dimensions of the system, but are independent of* n *and* ε.

This theorem, whose proof will be given in the next section, explains the excellent long-time behavior of the proposed variable step size algorithm for an important class of differential equations. It can further be extended to cover perturbed integrable systems. However, it cannot directly be applied to the Kepler problem (Figure 3.1), because there the frequencies do not satisfy the Diophantine condition. We expect that an argument similar to that of Example X.3.3 of [4] can be used for explaining the phenomena observed in this problem.

In many applications, particularly molecular dynamics simulation, one is concerned with nonintegrable Hamiltonian systems. Limited numerical experiments have shown that the adaptive Störmer–Verlet method (3.3) has an excellent long-time behavior (near conservation of the total energy) also in this situation.

4. Backward error analysis. We shall theoretically explain the excellent long-time behavior of the adaptive version (3.1) of reversible integrators. To this end, backward error analysis is an indispensable tool. Since the adaptive method (3.1) can be interpreted as a one-step method (3.2) in an augmented space, applied with a constant step size ε, standard arguments can be applied to obtain the modified equation. We assume that all appearing functions are real-analytic on suitable domains.

THEOREM 4.1. *Suppose that the basic method* Φ_h *is symmetric and reversible, and that* $GS = -G$. *Then there exist unique functions* $F_j(y, \rho)$ *and* $G_j(y, \rho)$ *such that, for an arbitrary truncation index* N, *the exact flow* $\widehat{\varphi}_t(y, \rho)$ *of the truncated modified equation*

$$(4.1) \qquad \begin{aligned} y' &= \left(F(y) + \varepsilon^2 F_1(y, \rho) + \cdots + \varepsilon^{2N-2} F_{N-1}(y, \rho) \right) / \rho, \\ \rho' &= G(y) + \varepsilon^2 G_1(y, \rho) + \cdots + \varepsilon^{2N-2} G_{N-1}(y, \rho) \end{aligned}$$

satisfies

$$(4.2) \qquad \widehat{\Phi}_\varepsilon(y_n, \rho_n) - \widehat{\varphi}_\varepsilon(y_n, \rho_n) = \mathrm{O}(\varepsilon^{2N+1}).$$

The expansions in (4.1) are in even powers of ε and the modified equation is reversible, i.e., $F_j \widehat{S} = -S F_j$ and $G_j \widehat{S} = -G_j$ for all j.

Proof. Inserting the solution of (4.1) into the method and comparing like powers of ε yields uniquely the functions F_j and G_j. The results of [4, sect. IX.2] then imply the reversibility of the modified equations and that the expansions are in even powers of ε.   □

If the basic method Φ_h is of order $2r > 2$, the leading perturbation terms in the series (4.1) are still of size $O(\varepsilon^2)$. This is because the interpolation of $\rho(\tau)$ is at the values ρ_n with integral indices and not at the values $\rho_{n+1/2}$. However, the coefficient functions satisfy

$$(4.3) \qquad F_j(y, \rho) = s_j(y, \rho) F(y) \qquad \text{for} \quad j = 1, \ldots, r-1,$$

with scalar functions s_j. This follows because Φ_h has a modified differential equation with leading perturbation term of size $O(h^{2r})$. The relation $y_{n+1} = \Phi_{\varepsilon/\rho_{n+1/2}}(y_n)$ therefore implies that the modified equation (4.1) for $y(\tau)$ has to be of the form $y' = s(y, \rho, \varepsilon) F(y) + O(\varepsilon^{2r})$ which is equivalent to (4.3).

4.1. Reversible perturbation theory. The numerical method (3.1) is consistent with the augmented system (2.12) as can be seen by putting $\varepsilon = 0$ in (4.1).

If $G(y)$ is given by (2.13) with Q satisfying $QS = Q$, the expression $A = Q(y)/\rho$ is a first integral of (2.12). Now assume that the original problem $\dot{y} = F(y)$ is an integrable reversible system and that $y = \psi(\theta, a)$ transforms it to action-angle variables. The transformation

$$(4.4) \qquad \begin{pmatrix} y \\ \rho \end{pmatrix} = \widehat{\psi}(\theta, a, A) = \begin{pmatrix} \psi(\theta, a) \\ Q(\psi(\theta, a))/A \end{pmatrix}$$

again preserves reversibility, i.e., $\widehat{S}\widehat{\psi} = \widehat{\psi}\widehat{S}$. It brings the system (2.12) to the form

$$(4.5) \qquad \theta' = r(\theta, a, A)\omega(a), \qquad a' = 0, \qquad A' = 0,$$

where $r(\theta, a, A) := A/Q(\psi(\theta, a))$. Here the number of action variables is increased by one. In contrast to (3.8), the differential equation for the angle variables depends not only on the actions, but also on the angles. This is characteristic of the present situation, because otherwise the step density ρ would be essentially constant, and we would not have an adaptive integrator.

By applying the coordinate change (4.4) to (4.1) we have

$$(4.6) \qquad \theta' = r(\theta, a, A)\Big(\omega(a) + \varepsilon^2 \eta_1(\theta, a, A) + \cdots + \varepsilon^{2N-2}\eta_{N-1}(\theta, a, A)\Big),$$

$$(4.7) \qquad a' = \varepsilon^2 \xi_1(\theta, a, A) + \varepsilon^4 \xi_2(\theta, a, A) + \cdots + \varepsilon^{2N-2}\xi_{N-1}(\theta, a, A),$$

$$(4.8) \qquad A' = \varepsilon^2 \zeta_1(\theta, a, A) + \varepsilon^4 \zeta_2(\theta, a, A) + \cdots + \varepsilon^{2N-2}\zeta_{N-1}(\theta, a, A),$$

which is an ε^2-perturbation of (4.5). Since the change of variables preserves reversibility, the functions r and η_j are even functions of θ, while the functions ξ_j and ζ_j are odd functions of θ.

This situation is similar to the one treated in [4, sect. XI.2], the only difference being that the unperturbed system depends on the angle variables. Since this dependence is only via the scalar function $r(\theta, a, A)$, the techniques and results of [4] can be extended to the present situation.

LEMMA 4.2. *Suppose that $\omega(b^*)$ satisfies the Diophantine condition (3.9). For a fixed $N \geq 2$, there then exists a reversibility-preserving change of coordinates*

$$(4.9) \qquad \theta = \varphi + \varepsilon^2 \mu_1(\varphi, b, B) + \cdots + \varepsilon^{2N-2}\mu_{N-1}(\varphi, b, B),$$

$$(4.10) \qquad a = b + \varepsilon^2 c_1(\varphi, b, B) + \cdots + \varepsilon^{2N-2}c_{N-1}(\varphi, b, B),$$

$$(4.11) \qquad A = B + \varepsilon^2 C_1(\varphi, b, B) + \cdots + \varepsilon^{2N-2}C_{N-1}(\varphi, b, B)$$

(i.e., μ_j is odd in φ, and c_j, C_j are even in φ) such that in the new coordinates, the system (4.6)–(4.8) is given by

$$(4.12) \qquad \varphi' = r(\varphi, b, B)\omega_{\varepsilon, N}(b, B) + \varepsilon^{2N} E_N(\varphi, b, B),$$

$$(4.13) \qquad b' = \varepsilon^{2N}\Xi_N(\varphi, b, B),$$

$$(4.14) \qquad B' = \varepsilon^{2N}Z_N(\varphi, b, B),$$

with $\omega_{\varepsilon, N}(b, B) = \omega(b) + \varepsilon^2 \omega_1(b, B) + \cdots + \varepsilon^{2N-2}\omega_{N-1}(b, B)$. The above transformation is defined for sufficiently small ε, for $\|b - b^*\| \leq c|\log \varepsilon|^{-\nu-1}$ with some ε-independent $c > 0$, for B in an ε-independent neighborhood of some B^*, and for φ in an ε-independent complex neighborhood of $\mathbb{T}^d$. The remainder terms in (4.12)–(4.14) are bounded independently of ε.

Proof. This proof is based on the ideas presented in sections X.2 and XI.2 of [4]. Here, we only focus on those parts which are different due to the dependence of the differential equation on the angle variables. We do not repeat technical details that can be taken over without changes.

Inserting (4.9)–(4.11) into the system (4.6)–(4.8) yields a differential equation for φ, b, and B. It is of the form (4.12)–(4.14) with $N = 2$ provided that the functions μ_1, c_1, C_1 satisfy

$$(4.15) \quad \frac{\partial}{\partial \varphi}\Big(\mu_1(\varphi, b, B)/r(\varphi, b, B)\Big)\omega(b) = \Big(\kappa(\varphi, b, B) - \omega_1(b, B)\Big)/r(\varphi, b, B),$$

$$(4.16) \qquad \frac{\partial}{\partial \varphi}\Big(c_1(\varphi, b, B)\Big)\omega(b) = \xi_1(\varphi, b, B)/r(\varphi, b, B),$$

$$(4.17) \qquad \frac{\partial}{\partial \varphi}\Big(C_1(\varphi, b, B)\Big)\omega(b) = \zeta_1(\varphi, b, B)/r(\varphi, b, B),$$

where $\kappa(\varphi, b, B)$ is an even function of φ depending on η_1, r, c_1, and C_1, but not on μ_1. To solve these equations for μ_1, c_1, and C_1, we represent all appearing functions as Fourier series, e.g.,

$$C_1(\varphi, b, B) = \sum_{k \in \mathbb{Z}^d}\gamma_k(b, B)e^{\mathrm{i}\,k\cdot\varphi}, \quad \zeta_1(\varphi, b, B)/r(\varphi, b, B) = \sum_{k \in \mathbb{Z}^d}\delta_k(b, B)e^{\mathrm{i}\,k\cdot\varphi}.$$

Equation (4.17) thus becomes

$$(4.18) \qquad \mathrm{i}\,k \cdot \omega(b)\,\gamma_k(b, B) = \delta_k(b, B).$$

Since the right-hand side of (4.17) is an odd function of φ, its angular average is zero, i.e., $\delta_0(b, B) = 0$, implying that (4.18) is automatically satisfied for $k = 0$. We can arbitrarily put $\gamma_0(b, B) = 0$. For $k \neq 0$ and $k \cdot \omega(b) \neq 0$, the relation (4.18) yields $\gamma_k(b, B)$. Assuming exponential decay of the Fourier coefficients δ_k and the Diophantine condition (3.9), the γ_k also decay exponentially, and $C_1(\varphi, b, B)$ is well-defined and a solution of (4.17). From $\delta_{-k} = -\delta_k$ it follows that $\gamma_{-k} = \gamma_k$, so

that the function C_1 is an even function of φ. Equation (4.16) can be solved in the same manner. Before solving (4.15), one has to define $\omega_1(b, B)$ such that the angular average of the right-hand side vanishes. □

For constant r the system obtained by neglecting the $\mathrm{O}(\varepsilon^{2N})$ terms in (4.12)–(4.14) can be solved exactly and yields constant functions $b(\tau)$ and $B(\tau)$ and a linear function for $\varphi(\tau)$. The next lemma shows that we have qualitatively the same behavior in the general case. We suppress the subscript in $\omega_{\varepsilon,N}(b, B)$ and the dependence on B (because b and B play the same role).

LEMMA 4.3. *Let $\omega(b_0)$ satisfy the Diophantine condition* (3.9), *and let $r(\varphi, b)$ be a scalar function which is analytic on a complex neighborhood of $\mathbb{T}^d \times \{b_0\}$ and satisfies $0 < r_0 \leq r(\varphi, b_0) \leq R_0$. The solution of the differential equation*

$$(4.19) \qquad\qquad \varphi' = r(\varphi, b_0)\omega(b_0)$$

then satisfies $\varphi(\tau) = \varphi_0 + \sigma(\tau)\omega(b_0)$, where $\sigma(\tau)$ is a monotonically increasing function with $\sigma(0) = 0$, and, for $\tau \to \infty$,

$$\sigma(\tau) = \mathrm{O}(\tau), \qquad \frac{\partial \sigma(\tau)}{\partial \varphi_0} = \mathrm{O}(1), \qquad \frac{\partial \sigma(\tau)}{\partial b_0} = \mathrm{O}(\tau).$$

Proof. Defining the time transformation $\sigma(\tau)$ as the solution of the differential equation $\sigma' = r(\varphi_0 + \sigma\omega(b_0), b_0)$, we see that $\varphi(\tau) := \varphi_0 + \sigma(\tau)\omega(b_0)$ solves (4.19). The boundedness of r implies that $\sigma(\tau) = \mathrm{O}(\tau)$. Separation of the variables in the differential equation for σ yields

$$(4.20) \qquad\qquad \int_0^{\sigma(\tau)} \frac{\mathrm{d}\sigma}{r(\varphi_0 + \sigma\omega(b_0), b_0)} = \tau,$$

and implicit differentiation with respect to φ_0 gives

$$(4.21) \quad \frac{\partial \sigma(\tau)}{\partial \varphi_0} \cdot \frac{1}{r(\varphi_0 + \sigma(\tau)\omega(b_0), b_0)} + \int_0^{\sigma(\tau)} \frac{\partial}{\partial \varphi_0}\Big(\frac{1}{r(\varphi_0 + \sigma\omega(b_0), b_0)}\Big)\mathrm{d}\sigma = 0.$$

Letting $\gamma_k(b_0)$ be the Fourier coefficients of the inverse of $r(\varphi, b_0)$, the integral in (4.21) becomes

$$\int_0^{\sigma(\tau)} \frac{\partial}{\partial \varphi_0}\Big(\sum_{k\in\mathbb{Z}^d} \gamma_k(b_0)e^{\mathrm{i}\,k\cdot(\varphi_0+\sigma\omega(b_0))}\Big)\mathrm{d}\sigma = \sum_{k\in\mathbb{Z}^d} \mathrm{i}\,k\,\gamma_k(b_0) \frac{e^{\mathrm{i}\,k\cdot(\varphi_0+\sigma\omega(b_0))}}{\mathrm{i}\,k\cdot\omega(b_0)}\Big|_0^{\sigma(\tau)}.$$

By the Diophantine condition and by the exponential decay of the Fourier coefficients this expression is bounded, implying that the derivative of $\sigma(\tau)$ with respect to φ_0 is bounded.

A similar argument with an additional partial integration gives the statement on the derivative with respect to b_0. □

4.2. Proof of Theorem 3.2. The numerical solution of our adaptive reversible algorithm is very close to the solution of the modified differential equation (4.1). To get more insight into this differential equation (and hence into the numerical solution) we have transformed it successively to the simpler form (4.12)–(4.14). We now have to transform properties of the system (4.12)–(4.14) back to those of the modified equations. This then proves Theorem 3.2.

(a) The exact solution of (4.12)–(4.14), with initial values φ_0, b_0, and B_0, satisfies

$$
\begin{aligned}
\varphi(\tau) &= \varphi_0 + \sigma(\tau)\omega_{\varepsilon,N}(b_0, B_0) + \mathrm{O}(\tau\varepsilon^{2N}) + \mathrm{O}(\tau^2\varepsilon^{2N}), \\
b(\tau) &= b_0 + \mathrm{O}(\tau\varepsilon^{2N}), \\
B(\tau) &= B_0 + \mathrm{O}(\tau\varepsilon^{2N}),
\end{aligned}
$$

(4.22)

where $\sigma(\tau)$ is the time transformation of Lemma 4.3. The statements for $b(\tau)$ and $B(\tau)$ follow from integration of (4.13) and (4.14). Inserting these estimates into (4.12) adds a term $\mathrm{O}(\tau\varepsilon^{2N})$ to the remainder. The nonlinear variation of constants formula together with the boundedness of $\partial\varphi/\partial\varphi_0$, which is a consequence of the estimates of Lemma 4.3, yields the formula for $\varphi(\tau)$.

(b) The transformation (4.9)–(4.11), which is ε^2-close to the identity, brings the statements (4.22) back to the action-angle variables θ, a, and A. As a consequence of the estimates of Lemma 4.3 we have

$$
\begin{aligned}
\theta(\tau) &= \theta_0 + \sigma(\tau)\omega_{\varepsilon,N}(a_0, A_0) + \mathrm{O}(\varepsilon^2) + \mathrm{O}(\tau\varepsilon^2) + \mathrm{O}(\tau^2\varepsilon^{2N}), \\
a(\tau) &= a_0 + \mathrm{O}(\varepsilon^2) + \mathrm{O}(\tau\varepsilon^{2N}), \\
A(\tau) &= A_0 + \mathrm{O}(\varepsilon^2) + \mathrm{O}(\tau\varepsilon^{2N}).
\end{aligned}
$$

(4.23)

The $\sigma(\tau)$ in this formula is given by (4.20) with b_0 replaced by a_0, A_0.

(c) Theorem 4.1 shows that the local error (difference between the numerical solution and the exact solution of the truncated modified differential equation) is bounded by $\mathrm{O}(\varepsilon^{2N+1})$. Since, for integrable systems, perturbations in the initial values are propagated at most linearly in time, this implies that the difference between the numerical solution and the exact solution of the modified equation (global error) increases at most as $\mathrm{O}(\tau^2\varepsilon^{2N})$ and that in the action variables at most as $\mathrm{O}(\tau\varepsilon^{2N})$. Together with (4.23) this proves that

$$
\begin{aligned}
I(p_n, q_n) &= I(p_0, q_0) + \mathrm{O}(\varepsilon^2) + \mathrm{O}(\tau\varepsilon^{2N}), \\
Q(p_n, q_n)/\rho_n &= Q(p_0, q_0)/\rho_0 + \mathrm{O}(\varepsilon^2) + \mathrm{O}(\tau\varepsilon^{2N})
\end{aligned}
$$

for the action variables in the augmented system. The error in the solution, which is essentially that in the angle variables, is

(4.24) $\qquad (p_n, q_n) - \big(p(\tau_n), q(\tau_n)\big) = \mathrm{O}(\varepsilon^2) + \mathrm{O}(\tau\varepsilon^2) + \mathrm{O}(\tau^2\varepsilon^{2N})$.

Here $\tau_n = n\varepsilon$, and $y(\tau) = (p(\tau), q(\tau))$ is the solution of the differential equation (2.12).

(d) To complete the proof of Theorem 3.2 we have to rewrite this estimate in the original time variable t. Adding $t_{n+1} = t_n + \varepsilon/\rho_{n+1/2}$ to (3.1) gives a symmetric discretization of (2.12) augmented by $t' = 1/\rho$. The backward error analysis of Theorem 4.1 then yields (4.1) augmented by a modified equation for $t(\tau)$:

(4.25) $\qquad t' = 1/\rho_\varepsilon(y, \rho), \qquad \rho_\varepsilon(y, \rho) = \rho + \varepsilon^2 r_1(y, \rho) + \cdots + \varepsilon^{2N-2} r_{N-1}(y, \rho)$.

(Putting the ε^2-series in the denominator does not change the argumentation.) The local error of the values t_n compared to the solution $\widehat{t}(\tau)$ of the modified equation (4.25) is $\mathrm{O}(\varepsilon^{2N+1})$, so that we have for the global error

(4.26) $\qquad\qquad\qquad t_n = \widehat{t}(\tau_n) + \mathrm{O}(\tau\varepsilon^{2N})$.

Since we have $t(\tau) - \widehat{t}(\tau) = \mathrm{O}(\tau\varepsilon^2)$ for the solution $t(\tau)$ of the unperturbed differential equation, we finally obtain $y(\tau_n) = y(t(\tau_n)) = y(\widehat{t}(\tau_n)) + \mathrm{O}(\tau\varepsilon^2) = y(t_n) + \mathrm{O}(\tau\varepsilon^2)$. Together with (4.24) this proves the statement of the theorem for order $2r = 2$. $\qquad\square$

(e) Slight modifications of the above proof are necessary to get the correct order of convergence for higher order methods. First of all we replace the factor $1/\rho$ in the modified differential equation (4.1) by $1/\rho_\varepsilon(y, \rho)$ with the function of (4.25). This implies that $F_j(y, \rho) = 0$ for $j = 1, \ldots, r-1$ as expected for a method of order $2r$. We then obtain (4.6)–(4.8) with a function $r(\theta, a, A)$ depending on ε^2, and $\eta_j = \xi_j = 0$ for $j = 1, \ldots, r - 1$. In Lemma 4.2 we then use a change of coordinates for which $\mu_j = c_j = 0$ for $j = 1, \ldots, r-1$. The function r in (4.12) is then $r(\theta, a, A)$ written in the new variables, and hence also depends on ε^2. Since the transformation (4.12)–(4.13) is ε^{2r}-close to the identity, the former proof yields the correct order of convergence.

5. Concluding remarks. We conclude this article with some remarks concerning further applications of the presented theory, and with comments related to a practical implementation.

5.1. Theoretical justification of proportional controllers. In the variable step size strategies of Hut, Makino, and McMillan [8] or Stoffer [11] the step size is defined implicitly by an algebraic relation of the type $e(y_n, y_{n+1}, h_{n+1/2}) = \varepsilon$, satisfying suitable symmetry and reversibility conditions. Backward error analysis [5] then yields a modified differential equation for y as in (4.1) with ρ replaced by a given function of y and ε. The resulting controller is *proportional* but not integrating, so there is no differential equation for the step density ρ. A simplified version of the above analysis (one need not consider the variables ρ, A, and B) then gives the same statement as in Theorem 3.2 for proportional controllers applied to integrable reversible systems.

5.2. Step size integrating controller. Instead of an integrating controller for the step density ρ as in (2.8), we consider a step size integrating controller (2.6). We put $h_{n+1/2} = \varepsilon s_{n+1/2}$, where $s_{n+1/2}$ are discrete values approximating a function that is given by a differential equation

$$(5.1) \qquad s' = H(y), \qquad s(0) = 1$$

(cf. (2.11)). Considering the control objective $Q(y)s = $ const, we obtain by differentiation with respect to τ and by using $y' = sF(y)$ that

$$(5.2) \qquad s' = -s^2 \nabla Q(y)^{\mathrm{T}} F(y)/Q(y) = -s^2 G(y),$$

with $G(y)$ from (2.13). Although not of the form (5.1), this differential equation can be solved for constant $y = y_n$ by separation of variables and yields

$$\frac{1}{s_{n+1/2}} - \frac{1}{s_{n-1/2}} = \varepsilon\, G(y_n),$$

which again is the integrating controller for the step density $\rho = 1/s$.

Insisting on an integrating controller for the step size, we can replace s in (5.2) by $s = \mathrm{const}/Q(y) = Q(y_0)/Q(y)$ to obtain

$$(5.3) \qquad s' = -Q(y_0)^2 \nabla Q(y)^{\mathrm{T}} F(y)/Q(y)^3,$$

which then leads to

$$s_{n+1/2} = s_{n-1/2} - \varepsilon\, Q(y_0)^2 \nabla Q(y_n)^{\mathrm{T}} F(y_n)/Q(y_n)^3,$$

with the step size given by $h_{n+1/2} = \varepsilon s_{n+1/2}$. Numerical experiments indicate that this controller also performs well, without drift in the numerical energy, when applied to the problem of Figure 3.1.

However, the theory of the previous section no longer applies. Since the differential equation (5.3) depends on the initial value y_0, the expression $Q(y)s$ is no longer a first integral of the system $y' = sF(y)$ augmented by (5.3), because

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\Big(Q(y)\,s\Big) = \nabla Q(y)F(y)\left(s^2 - \frac{Q(y_0)^2}{Q(y)^2}\right),$$

which vanishes only on the manifold defined by $Q(y)s = Q(y_0)$. Note that $Q(y)s$ is a first integral of (5.2). This lack of theoretical justification is one of the main reasons for proposing the integrating controller for the step density ρ and not for the step size s.

5.3. Implementation issues. Given a control function $G(y)$, the reversible step density controller is implemented in the form

(5.4) $$\rho_{n+1/2} = \rho_{n-1/2} + \varepsilon\,\alpha\,G(y_n),$$

with initial value $\rho_0 = 1$ and $\rho_{1/2} = \rho_0 + \varepsilon\,\alpha\,G(y_0)/\varepsilon$. The controller has two settings. The *setpoint* $\varepsilon > 0$ is the external means of controlling *accuracy*, via the step size $h_{n+1/2} = \varepsilon/\rho_{n+1/2}$. It affects the *absolute magnitude* of the entire step size sequence.

The controller's *integral gain* $\alpha \geq 0$ (see [9] for control theoretic terminology) is used to balance the *computational effort*, by affecting the rate of change in the step density. This parameter was used in (3.6). Adjusting the integral gain is equivalent to changing the control objective; with gain α, the objective is $Q(y)^\alpha/\rho = \text{const}$. Taking $\alpha = 0$ produces constant step size, while $\alpha = 1$ produces the controller discussed in previous sections.

REFERENCES

[1] M. P. Calvo and J. M. Sanz-Serna, *Variable steps for symplectic integrators*, in Numerical Analysis 1991, Pitman, Res. Notes Math. Ser. 260, Longman, Harlow, UK, 1992, pp. 34–48.

[2] S. Cirilli, E. Hairer, and B. Leimkuhler, *Asymptotic error analysis of the adaptive Verlet method*, BIT, 39 (1999), pp. 25–33.

[3] B. Gladman, M. Duncan, and J. Candy, *Symplectic integrators for long-term integrations in celestial mechanics*, Celest. Mech. Dyn. Astr., 52 (1991), pp. 221–240.

[4] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*, Springer Ser. Comput. Math. 31, Springer, Berlin, 2002.

[5] E. Hairer and D. Stoffer, *Reversible long-term integration with variable stepsizes*, SIAM J. Sci. Comput., 18 (1997), pp. 257–269.

[6] T. Holder, B. Leimkuhler, and S. Reich, *Explicit variable step-size and time-reversible integration*, Appl. Numer. Math., 39 (2001), pp. 367–377.

[7] W. Huang and B. Leimkuhler, *The adaptive Verlet method*, SIAM J. Sci. Comput., 18 (1997), pp. 239–256.

[8] P. Hut, J. Makino, and S. McMillan, *Building a better leapfrog*, Astrophys. J., 443 (1995), pp. 93–96.

[9] G. Söderlind, *Automatic control and adaptive time–stepping*, Numer. Algorithms, 31 (2002), pp. 281–310.

[10] G. Söderlind, *Digital filters in adaptive time–stepping*, ACM Trans. Math. Software, 29 (2003), pp. 1–26.

[11] D. Stoffer, *Variable steps for reversible integration methods*, Computing, 55 (1995), pp. 1–22.