SIAM J. SCI. COMPUT.
Vol. 20, No. 3, pp. 930–963

LONG-TIME-STEP METHODS FOR OSCILLATORY DIFFERENTIAL EQUATIONS*

B. GARCÍA-ARCHILLA†, J. M. SANZ-SERNA‡, AND R. D. SKEEL§

Abstract. Considered are numerical integration schemes for nondissipative dynamical systems in which multiple time scales are present. It is assumed that one can do an explicit separation of the RHS "forces" into fast forces and slow forces such that (i) the fast forces contain the high frequency part of the solution, (ii) the fast forces are conservative, and (iii) the reduced problem consisting only of the fast forces can be integrated much more cheaply than the full problem. The fast forces are allowed to have low frequency components. Particular applications for which the schemes are intended include N-body problems (for which most of the forces are slow) and nonlinear wave phenomena (for which the fast forces can be propagated by spectral methods). The assumption of cheap integration of fast forces implies that the overall cost of integration is primarily determined by the step size used to sample the slow forces. A long-time-step method is one in which this step size exceeds half the period of the fastest normal mode present in the full system. An existing method that comes close to qualifying is the "impulse" method, also known as Verlet-I and r-RESPA. It is shown that it might fail, though, for a couple of reasons. First, it suffers a serious loss of accuracy if the step size is near a multiple of the period of a normal mode, and, second, it is unstable if the step size is near a multiple of half the period of a normal mode. Proposed in this paper is a "mollified" impulse method having an error bound that is independent of the frequency of the fast forces. It is also shown to possess superior stability properties. Theoretical results are supplemented by numerical experiments. The method is efficient and reasonably easy to implement.

Key words. Verlet method, multiple time steps, oscillatory differential equations, long time steps, r-RESPA

AMS subject classifications. 65L05, 65M05, 70F10

PII. S1064827596313851

1. Introduction.

We consider the numerical integration of systems of special second-order differential equations with multiple time scales such as those arising from N-body problems and from the spatial discretization of partial differential equations describing wave phenomena. More specifically, we assume that the system has the form

$$M\frac{d^2}{dt^2}q = -W_q(q) + F(q),$$

where M is a diagonal mass matrix and the RHS force vector is explicitly split into two parts, the first part being a gradient and contributing fast modes to the motion and the second part not containing any fast modes. More precisely, the first term is fast in the sense that $M^{-1/2}W_{qq}M^{-1/2}$ has some large positive eigenvalues, and the

*Received by the editors December 18, 1996; accepted for publication (in revised form) June 15, 1997; published electronically October 20, 1998.
http://www.siam.org/journals/sisc/20-3/31385.html

†Departamento de Matemáticas, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain (bosco@ccuam3.sdi.uam.es). This research was supported in part by grant DGICYT PB95-216.

‡Departamento de Matemática Aplicada y Computación, Facultad de Ciencias, Universidad de Valladolid, Valladolid, Spain (sanzserna@cpd.uva.es). This research was supported in part by grant DGICYT PB95-705.

§Department of Computer Science (and Beckman Institute), University of Illinois at Urbana-Champaign, Urbana, IL 61801-2987 (skeel@cs.uiuc.edu). This research was performed mostly while at the University of Valladolid as Iberdrola Visiting Professor, with additional support from NSF grants DMS-9600088 and BIR-9318159 and NIH grant P41RR05969.

second term is slow in the sense that $-M^{-1/2}F_qM^{-1/2}$ has only small eigenvalues. The first term is allowed to have eigenvalues of small modulus. If $F(q)$ is a gradient $-U_q(q)$, our system is a Hamiltonian system and we ask then that our integrator be symplectic [12].

In special cases where $W(q)$ has the form $\frac{1}{\epsilon}\sum_j g_j(q)^2$, it is possible to consider imposing constraints $g_i(q) = 0$ thus obtaining *reduced variable dynamics*. This can be justified by averaging arguments, which often involve the addition of terms to the differential equation in order to account for things that do not average out to zero, e.g., [3]. The computational costs of computing these additional terms and of imposing the constraints can be considerable.

We study in this paper a less radical approach based on approximation of the *full dynamics*. Many of the interactions that constitute the collection of forces in a given physical problem can be permanently classified as fast or slow. For interactions of variable speed, it may be computationally efficient to split them artificially into fast and slow parts [14]. The idea behind the division into $-W_q(q) + F(q)$ is to sample terms of $F(q)$ infrequently and incorporate them into a reduced problem involving $W(q)$. We define a *long-time-step method* to be one that samples the slow force at time increments greater than half the period[1] of the fastest oscillation in the system. The reduced problem might be solved analytically or it might be solved by a numerical scheme using shorter step sizes, in which case the overall method is a two-time-step scheme. This idea might, of course, be applied recursively—the reduced problem itself can be solved with a not-quite-as-long-time-step method, and so on, resulting in a hierarchy of step sizes. The fast/slow splitting is worthwhile only if the cost of integrating the reduced system is much less than the cost of integrating the whole system. This could happen for either or both of the following reasons.

1. The bulk of the force calculation involves interactions that are slow, which is the case for molecular dynamics and other N-body problems. Also, it might be noted that the fast interactions tend to be local in space so that in a parallel implementation, slow interactions are likely to be those that require the most communication.
2. The cost of integrating the reduced system for long step sizes is not much greater than the cost for short step sizes. This is the case if the reduced system can be solved analytically, for example, using spectral methods.

One potential candidate for long-time-step integration is the Verlet-I/r-RESPA impulse method [7, 15], but in practice it seems to qualify only as a medium-time-step method. In this paper, we explain the poor behavior of the impulse method by an analysis of its stability and accuracy. We also propose a nontrivial improvement of the impulse method that we call the *mollified impulse method*, for which superior stability and accuracy is demonstrated. Numerical evidence is also provided.

The impulse method can be expressed using the Dirac delta function as the following approximation:

$$M\frac{d^2}{dt^2}q = -W_q(q) + \sum_{n=-\infty}^{+\infty} h\delta(t - nh)F(q), \tag{1}$$

a formulation first published in [17]. Hence in the so-called endpoint version, a step $n \to n + 1$ of the method can be described, with $p = M(d/dt)q$, as follows:

[1] This seems to represent a theoretical barrier of some kind.

kick add $(h/2)F(q_n)$ to p_n to get p_n^+;

oscillate use the h-flow of $(d/dt)p = -W_q(q)$, $(d/dt)q = M^{-1}p$ to advance from (p_n^+, q_n) to (p_{n+1}^-, q_{n+1});

kick add $(h/2)F(q_{n+1})$ to p_{n+1}^- to get p_{n+1}.

Note that the force $F(q^{n+1})$ at the second kick of the current step coincides with the force at the first kick of the next step; hence there is really one impulse at each step point $t = nh$. The impulse method is derived as a multiple-time-step (MTS) method in [5, 7], but these writings express little enthusiasm for the method because of the possibility of resonance if the period h of the impulse should happen to coincide with a natural frequency of the reduced system $M(d^2/dt^2)q = -W_q(q)$. The resonance is demonstrated experimentally in [1]. Also, molecular dynamics experiments in [4] seem to indicate that the step size has to be less than the resonance value, which is 9–10 fs for fully flexible classical mechanics models of molecules. Other experiments [6, 8] show the inferiority of the impulse method (Verlet-I) in a Langevin dynamics setting, in which a random noise term and a balancing damping term are added to Newton's second law of motion. The restriction $h\Omega < 2\pi$ is assumed in [18] for all frequencies Ω present in a Poisson series. In sections 2 and 6 of this paper it is shown that the impulse method is not uniformly convergent—it undergoes an order reduction from two to one in positions and from two to zero in velocities as h approaches $2\pi/\Omega$. A much smaller limit on the step size, however, is suggested by other computational experience with the impulse method [9, 16], in particular, a step size of less than 5 fs is needed in molecular dynamics to prevent energy growth. Moreover, recent experiments on a large molecular system [2] show that for step sizes of 6 to 7 fs the energy growth is less severe than for 5 fs. We give in this paper a linear analysis that reveals instability for step sizes just less than half the shortest period of any normal mode. This instability can produce exponential energy growth as time increases, regardless of how soft the slow forces are. In technical terms, the impulse MTS method is only "weakly" stable for long step sizes. In conclusion the impulse method is not a long-time-step method in the sense we defined above; i.e. it cannot operate successfully when sampling the slow force at time increments greater than half the period of the fastest oscillations in the problem.

It is very important to emphasize that we assume that, in the impulse method, the reduced problem $(d/dt)p = -W_q(q)$, $(d/dt)q = M^{-1}p$ is solved either analytically or with "short" step sizes. This implies that the method is keeping track of all the fast scales present in the problem and, in particular, would be *exact* in the absence of slow forces. Its shortcomings may therefore come as an unpleasant surprise.

The enhancement that we propose for the impulse method involves modifying the strength of the impulses $F(q_n)$ to become $\mathcal{A}_q(h; q_n)^{\mathrm{T}} F(\mathcal{A}(h; q_n))$. Evaluation of F at $\mathcal{A}(h; q_n)$ represents an averaging; $F(\mathcal{A}(h; q_n))$ is expected to be a better description of the quickly varying $F(q(t))$ than the values of F at step points used by the impulse method. The transpose Jacobian $\mathcal{A}_q(h; q_n)^{\mathrm{T}}$ compensates for the effect of treating the slow force as an impulse (see the opening example in section 2). Different choices of the averaging operator $\mathcal{A}(h; q_n)$ are possible and lead to different numerical methods.

As is the case in the impulse method, the method suggested here uses an "exact" integration of the reduced problem $(d/dt)p = -W_q(q)$, $(d/dt)q = M^{-1}p$ and would be exact in the absence of slow forces. This is at variance with the situation for other conceivable methods that would use *averaging* of the reduced problem in order to avoid keeping track of the fast oscillations. In the method suggested here averaging is only performed to incorporate the slow forces into the (unaveraged) fast dynamics.

Numerical tests confirm that the modification to the impulse method suggested here yields a dramatic improvement. For the modified methods we obtain a bound on the "global error"—the error after many long steps on a predetermined time interval. Under the assumption that $W(q)$ is a semipositive definite quadratic form, we prove second-order accuracy for positions and first-order accuracy for momenta, with error bounds that depend only on the "reduced" energy $\frac{1}{2}p^{\mathrm{T}}M^{-1}p+W(q)$ and derivatives of $F(q)$ and *not on derivatives of* $W(q)$. The independence of the bounds on derivatives of $W(q)$ implies that, provided that the reduced energy is kept bounded, it is possible to apply the method with a given time-step h to faster and faster problems without impairing the accuracy. A bound for p and q in terms of the energy implies less *relative* accuracy for higher frequencies, because, for a given value of energy, the amplitudes of high frequency modes must get closer to zero as the frequency gets higher. For frequencies of order $O(h^{-1})$ only first-order relative accuracy is attained by the suggested methods, and for frequencies of order $O(h^{-2})$ or greater no relative accuracy is attained. Hence for high enough frequencies, the suggested methods do not resolve their contributions. If such frequencies are present, we might call the problem "stiff-oscillatory"; the suggested methods are then stiff-oscillatory solvers in the sense that they only resolve the oscillations that contribute with significant amplitudes.

Section 3 considers fixed-h stability as $t \to \infty$. If W_{qq} were absent, the method would reduce to the Störmer/leapfrog/Verlet method and stability would be ensured for $h\omega$ less than 2, where ω^2 is the spectral radius of $-F_q$. Ideally, this same stability restriction would still suffice with h^2W_{qq} present. A stability analysis is presented for a problem with two degrees of freedom where both W_{qq} and $-F_q$ are symmetric semipositive definite matrices. We study for various averaging operators how the stability depends on $h\Omega_1$, $h\Omega_2$, and $h\omega$ where Ω_1^2 and Ω_2^2 are the eigenvalues of W_{qq}. We discover that ideal stability is achieved only for rather special averagings, which are generally not easy to implement. The generally more practical averagings suggested in the next section all exhibit instabilities in at least some—very narrow—regions of parameter space even for arbitrarily small $h\omega$. The simple impulse method has an especially large number of instability regions, whereas the method we call "LongAverage" is unstable in considerably fewer situations. LongAverage has instabilities when the sum $h\Omega_1 + h\Omega_2$ is approximately some positive integer multiple of 2π. The instability is present only if the slow force creates a coupling between the two high frequencies. A cheaper method we call "ShortAverage" and a method we call "LinearAverage" suffer for an instability similar to that of LongAverage and are also unstable when $h\Omega_1$ or $h\Omega_2$ is near some odd multiple of π. The hapless (unmodified) impulse method and ShortAverage suffer from the instabilities of Short- and LinearAverage and are additionally unstable when either $h\Omega_1$ or $h\Omega_2$ is near some positive integer multiple of 2π. Figures 1–5 of section 3 display the dangerous regions of parameter space for $h\omega = \frac{1}{2}$. For *medium* step sizes LongAverage offers an advantage over the unmodified impulse method. More specifically, LongAverage is stable in the octant of parameter space defined by $h\Omega_i < \pi$ and $h\omega < 2$, whereas the simple impulse method and the two other proposed averagings are not. (The simple impulse method is unstable[2] for $h\Omega_i = \frac{3}{4}\pi$ and $h\omega = \sqrt{2}$.) For nonlinear problems, for which the analysis does not apply, we test numerically how effective these methods are and observe the effect of nonlinearity in ameliorating the stability problems. Nonlinearity helps to stabilize but not always enough. However, slight changes in the step size can effect a dramatic

[2]Choose $\alpha_i = \frac{1}{2}\omega^2$ and $\beta = -\frac{1}{2}\omega^2$ in section 7 and thus violate condition 2 of Lemma 5.

improvement. In many practical situations it should be possible to avoid instabilities through a knowledge of the values of the high frequencies present in the problem.

In molecular dynamics, $W(q)$ corresponds to bonded interactions and possibly the "short-range parts" of nonbonded interactions. These are by no means quadratic; however, on a time interval in which the system stays within the basin of a local minimum of $W(q)$, there is a symplectic change of variables such that $W(q)$ is nearly quadratic. Moreover, there is empirical evidence [19] of harmonic behavior over time scales approaching several hundred femtoseconds.

The paper has been organized so that the information required to use the methods is presented first and the more theoretical material appears toward the end. The sections are as follows:

2. description of the algorithm,
3. stability for fixed step size as $t \to \infty$,
4. accuracy,
5. derivation of the method,
6. error bounds and convergence, and
7. stability analysis.

2. The mollified impulse method. In the impulse method and its improvements, there are two natural ways to formulate the algorithm:

1. midpoint version: oscillate half way; kick; oscillate half way.
2. endpoint version: kick half way; oscillate; kick half way.

We focus on the latter, which has two advantages.

1. *Computational efficiency.* If the endpoint idea is applied recursively, it results in an algorithm in which all faster forces are being computed whenever slower forces are computed. In cases where a fast and a slow force are created artificially by splitting a force of variable speed, only a little extra work is needed to get both the fast and the slow part of such force. Also in a parallel message-passing computational environment, the number of messages communicating positional coordinates is reduced.
2. *Convenience.* The simplest schemes for attempting to interpolate slow forces require data only at the two endpoints.

As pointed out in the introduction, the suggested method is obtained from the simple impulse method, described after equation (1), by replacing the impulses $F(q_n)$ by $\mathcal{A}_q(h;q_n)^{\mathrm{T}}F(\mathcal{A}(h;q_n))$. By means of a simple example, we explain why we need to mollify the force by multiplying by $\mathcal{A}_q(h;q_n)^{\mathrm{T}}$. The differential equation

$$\frac{d^2}{dt^2}q = -\Omega^2 q + F,$$

where $\Omega \gg 1$, describes the displacement of a unit mass subject to the pull of a stiff spring held fixed at the other end and to the pull of a constant force F. For simplicity assume initial values $q(0) = 0$ and $p(0) = 1$. Numerical integration by the impulse method with step size h incorporates the slow force F by adding a term $\frac{1}{2}hF$ to the momentum at the beginning and at the end of every step. Suppose though that h has been chosen so that $h\Omega = 2\pi$. Between the impulses the reduced problem $(d^2/dt^2)q = -\Omega^2 q$ is integrated exactly; and because any solution of this problem has period $h = 2\pi/\Omega$, the result of integrating will be to leave the value of p and q exactly unchanged. Hence each complete step adds hF to p and leaves q unchanged. This happens to be correct for q but *not at all* for p. The correct solution is

$$p(t) = \cos t\Omega \cdot 1 + \Omega^{-1} \sin t\Omega \cdot F \tag{2}$$

and has the constant value 1 at integer multiples nh of h. Equation (2) shows that, in the true dynamics, the effect F is mollified by multiplication by $\Omega^{-1}\sin t\Omega$.

The action of $\mathcal{A}_q(h;q_n)^{\mathrm{T}}$ on the force can be seen not only as a mollification but also as a filter damping some components of the force. Different averaging procedures give rise to different filters. This point of view is taken up in section 5.

In order to find the average $\mathcal{A}(h;q_n)$, we use an interpolation, and there is flexibility in how this is done. Let ϕ be a basis function for interpolation on a mesh consisting of all integers so that $\sum_n \phi((t-nh)/h)g_n$ is the interpolant of data g_n on a mesh of spacing h. Consistency requires that

$$\sum_n \phi(s-n) \equiv 1,$$

which, by standard Fourier analysis techniques, can be easily shown to imply

$$\int_{-\infty}^{+\infty} \phi(s)\,ds = 1. \tag{3}$$

There are three interesting simple choices of ϕ:

1. the ShortAverage $\quad \phi(s) = \begin{cases} 1, & |s| < \frac{1}{2}, \\ \frac{1}{2}, & |s| = \frac{1}{2}, \\ 0, & |s| > \frac{1}{2}; \end{cases}$
2. the LongAverage $\quad \phi(s) = \begin{cases} \frac{1}{2}, & |s| < 1, \\ \frac{1}{4}, & |s| = 1, \\ 0, & |s| > 1; \end{cases}$
3. the LinearAverage $\quad \phi(s) = \begin{cases} 1-|s|, & |s| \le 1, \\ 0, & |s| \ge 1. \end{cases}$

The average $\mathcal{A}(h;q_n)$ is defined in terms of the solution $p(t;q_n)$, $q(t;q_n)$, $b(t;q_n)$ of an auxiliary initial value problem

$$\frac{d}{dt}p = -W_q(q), \quad \frac{d}{dt}q = M^{-1}p, \quad \frac{d}{dt}b = \phi\left(\frac{t}{h}\right)q,$$

with initial conditions $p(0)=0$, $q(0)=q_n$, $b(0)=0$. Note that the initial momentum is zero and that only fast forces are integrated. We define

$$\mathcal{A}(h;q_n) = \frac{1}{h}(b(+\infty;q_n) - b(-\infty;q_n))$$

so that

$$\mathcal{A}(h;q_n) = \frac{1}{h}\int_{-\infty}^{\infty} \phi\left(\frac{t}{h}\right) q(t)\,dt = \int_{-\infty}^{\infty} \phi(s)q(hs)\,ds.$$

In Short-, Long-, and LinearAverage and in other potentially useful choices, ϕ is an even function and the average reduces to $\frac{2}{h}b(+\infty;q_n)$. If, furthermore, $\phi(s)$ vanishes for $|s|>\mu$, then the average is simply $\frac{2}{h}b(\mu h;q_n)$ and the integration of the auxiliary problem is only required on a bounded t-interval. An infinite integration may be possible if the auxiliary problem can be integrated analytically in closed form. Otherwise, ϕ should be chosen to have bounded support.

The mollified method also requires the Jacobian matrix $\mathcal{A}_q(h;q_n) = \frac{2}{h}b_q(\mu h;q_n)$ formed by differentiating $\frac{2}{h}b(\mu h;q_n)$ with respect to q_n. This means that at the same

time we compute the average we have to compute derivatives of the average. We emphasize that, at each value of n, the auxiliary integration is only used to compute the mollified impulse. Once the impulse has been added to the momentum, the averaged position and its Jacobian matrix are discarded and the main integration is continued from q_n, which has not changed at all during the auxiliary integration.

To solve the auxiliary averaging problem, we should use exactly the same method, analytical or numerical, as that used to integrate the reduced problem (between evaluations of the slow force). For example, suppose that we are using the Verlet/leapfrog/Störmer method with a small time step δt. Then the calculation of the average and its Jacobian matrix should be done as follows:

initially:

$p := 0;$	$p_q := 0;$
$q := q_n;$	$q_q := I;$
$b := 0;$	$b_q := 0;$
$t := 0;$	

step by step:

$p := p - \frac{1}{2}\delta t W_q(q);$	$p_q := p_q - \frac{1}{2}\delta t W_{qq}(q)q_q;$
$b := b + \frac{1}{2}\delta t \phi(t/h)q;$	$b_q := b_q + \frac{1}{2}\delta t \phi(t/h)q_q;$
$q := q + \delta t M^{-1} p;$	$q_q := q_q + \delta t M^{-1} p_q;$
$t := t + \delta t;$	
$p := p - \frac{1}{2}\delta t W_q(q);$	$p_q := p_q - \frac{1}{2}\delta t W_{qq}(q)q_q;$
$b := b + \frac{1}{2}\delta t \phi(t/h)q;$	$b_q := b_q + \frac{1}{2}\delta t \phi(t/h)q_q.$

We compute step by step until we reach a value of t such that $\phi(t/h)$ is zero at this value and remains zero for larger values of t. For example, for LongAverage this means getting the value $b(h + \frac{1}{2}\delta t)$ because $\phi(t/h)$ vanishes only for $t > h$. (Equivalently, for the purpose of programming, we can define $\phi(1)$ to be $\frac{1}{2}$ rather than $\frac{1}{4}$ and stop at $t = h$.) Note that in this setting ShortAverage is cheaper than both Long- and LinearAverage. Note also that we compute derivatives of the numerical solution rather than numerically solving the variational problem.

Evaluating $W_{qq}(q)q_q$ is not as difficult as it might seem. One needs to make fairly generous use of the chain rule. For example, if we can write a 2-body interaction as

$$\frac{1}{2}\chi(\|\vec{r}_2 - \vec{r}_1\|^2),$$

then we can separately program (i) the first and second derivatives of the scalar function χ and (ii) the gradient and Hessian of $\|\vec{r}_2 - \vec{r}_1\|^2$. For additional details see [10].

An immediate question of practical (and theoretical) interest is, to what extent is it necessary to do the averaging integration as prescribed in this paper? Would it be almost as effective if, for example,

- the step size for the averaging integration was double that used for the reduced problem?
- the averaging integration used only the fastest part of $-W_q$?
- the averaging integration used a lower order of accuracy?

3. Stability. We study numerical stability for a system of three springs in two-dimensional physical space: the first spring of stiffness Ω_1^2 is fixed at the left at $(0, 0)$ and is attached at the right to a unit mass, the second spring of stiffness $\frac{1}{2}$ joins the

first unit mass to a second unit mass, the third spring of stiffness Ω_2^2 is attached at the left to the second unit mass and fixed at the right at $(0, 3)$. All three springs have unit length in the absence of forces. The choice of the stiffness of the slow spring is such that, when $\Omega_1 = \Omega_2 = 0$, the system has an angular frequency $\omega = 1$ leading to a stability restriction $h = h\omega < 2$ for the Verlet method. In molecular dynamics, typical values for the step size in the Verlet method are those for which h times the fastest frequency is in the range 1/6 to 2/3 (which corresponds to a step size in the range 0.25–1 fs).

The problem is nonlinear with four degrees of freedom. However, if the initial displacements and momenta have no vertical components, then the problem is linear and only has two degrees of freedom. In this linear case, the only considered until further notice, stability can be investigated by using the results of section 7. If the positions of the two masses are expressed as $(1 + q_1, 0)$ and $(2 + q_2, 0)$, the dynamics are given by (27) with $\alpha_i = \frac{1}{2}$ and $\beta = -\frac{1}{2}$. According to Lemma 4, stability of the numerical solution depends on the three parameters $\Gamma_1 = h\Omega_1$, h, and $\Gamma_2 = h\Omega_2$. Probably the most useful way to visualize matters is to plot for given values of h the boundaries of those regions in Γ_1–Γ_2 parameter space for which the numerical solution is stable.

Stability is determined by the four eigenvalues of the propagation matrix. These eigenvalues approximate the two frequencies present in the system. For small enough step sizes h, the eigenvalues are two complex conjugate pairs with unit modulus. As h increases, the eigenvalues can coalesce and move off the unit circle of the complex plane. Hence boundaries of stability regions are characterized by eigenvalues having multiplicity greater than 1. Three possibilities for instability are identified by Lemma 5:

type 1 instability is caused by the coalescing of imaginary eigenvalues at a point $\neq \pm 1$ and their subsequent movement off of the unit circle. This can occur when the sum $h\Omega_1 + h\Omega_2$ is near some positive integer multiple of 2π.

type 2 instability is caused by the coalescing of imaginary eigenvalues at -1. This can occur when either $h\Omega_1$ or $h\Omega_2$ is near some odd multiple of π.

type 3 instability is caused by the coalescing of imaginary eigenvalues at $+1$. This can occur when either $h\Omega_1$ or $h\Omega_2$ is near some positive integer multiple of 2π.

Boundaries of the instability regions satisfy (29)–(31) of section 7. These equations can be used to plot instability regions. Our plots are for the typical value $h = \frac{1}{2}$.

It is shown in section 7 that regions of type-1 instability are present for all four methods (impulse, LongAverage, ShortAverage, and LinearAverage). However, LongAverage is the only one of the four not exhibiting other types of instability. Hence it is attractive as a practical method, and it is interesting to examine its instabilities. The most important instability region in the Γ_1–Γ_2 plane is the one near the line from $(2\pi, 0)$ to $(0, 2\pi)$. It is the first that we encounter as we increase the step size with given Ω_1, Ω_2 and is the widest; it is shown in Figure 1 for $\Gamma_1 > \Gamma_2$ (the region is symmetric in Γ_1, Γ_2). Figures 2 and 3 show closeups from one end of the region near $\Gamma_2 = 0$ where it is a strip of width approximately 0.1 and from the other end near (π, π) where the strip has a width of around 3×10^{-7}. Figure 4 is a three-dimensional plot of the spectral radius of the propagation matrix.

A three-dimensional plot for the spectral radius of the ShortAverage propagation matrix is given in Figure 5. The type-2 instability region shows up well, but the mesh used for plotting is too coarse to reveal the type-1 instability region.

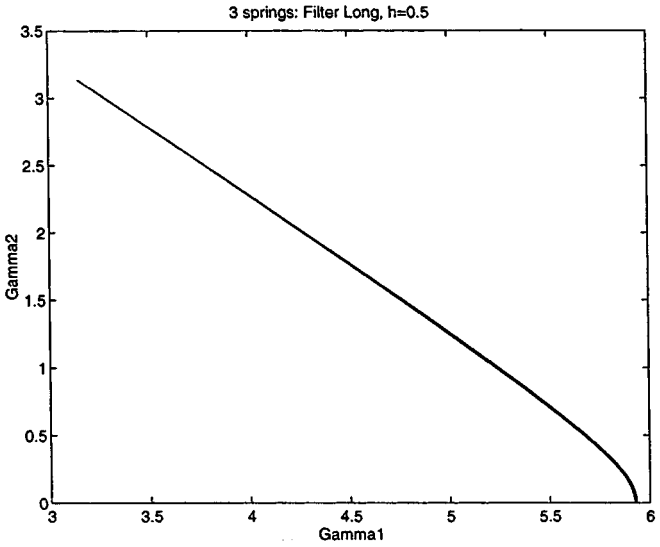

FIG. 1. *Instability region (strip) of LongAverage method.*

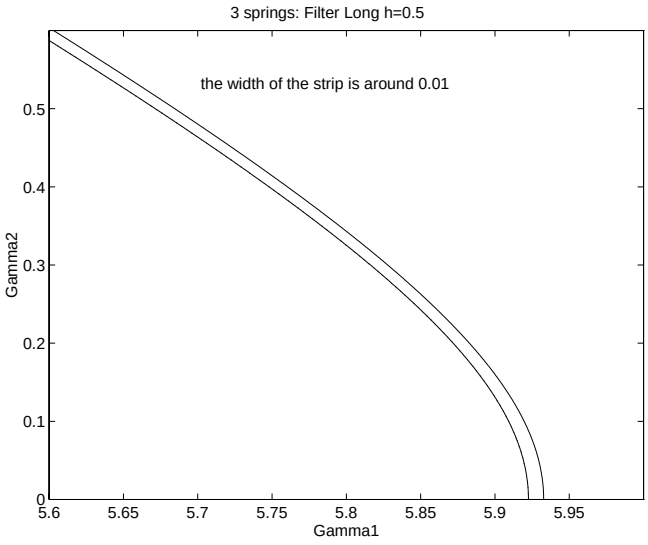

FIG. 2. *Closeup of instability strip of LongAverage at wide end.*

Not unexpectedly, one can increase the step size further for LongAverage than for the other methods before the first instability region is encountered. It is shown in section 7 that for the Γ_i less than π, the stability threshold for the $h\omega_i$ is uniformly 2. For the other methods as the Γ_i approaches π, the stability threshold for the $h\omega_i$ goes to 0.

How might we, in practice, cope with the possibility of instability? For molecular dynamics it is worthwhile to customize numerical methods because these simulations often run for months and because the properties of molecules do not change. Hence it seems reasonable to determine the high frequencies and then choose h to avoid regions of stability. We also can and should monitor the simulation to guard against instabilities. For applications where spectral methods are being used, it follows from

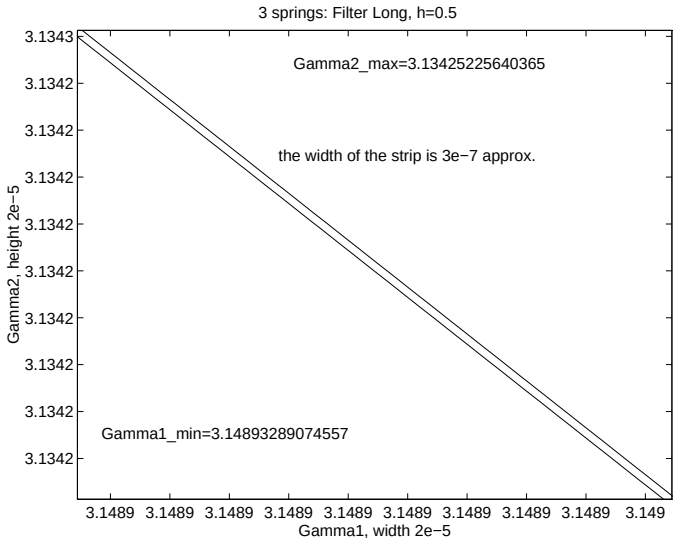

FIG. 3. *Closeup of instability strip of LongAverage at narrow end.*

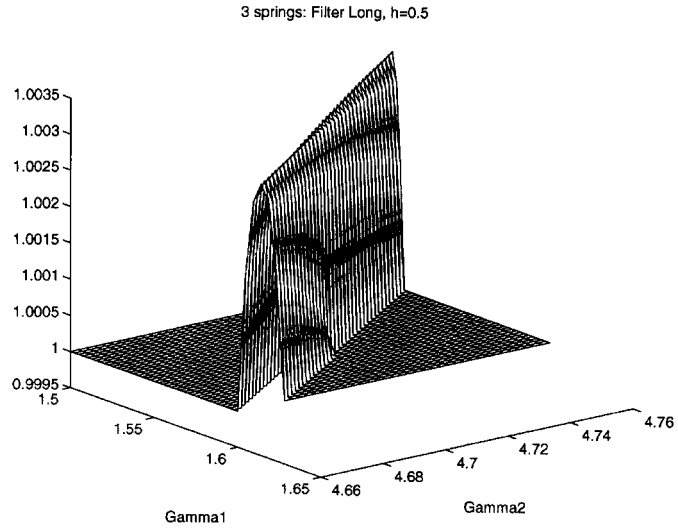

FIG. 4. *Spectral radius of the propagation matrix for LongAverage method.*

the analysis of type-1 instability in section 7 that we can easily create a method without regions of instability.

We supplement the study of the stability regions with numerical experiments to get some idea of the strength and robustness of the instabilities. In all simulations in this paper, the reduced problem and the auxiliary problems are integrated by the leapfrog–Verlet method as described in section 2. We do long-time integrations with $h = \frac{1}{2}$ with the impulse method, the (more stable) LongAverage method, and the (cheaper) ShortAverage method. We select three different sets of stiffnesses chosen to give each method a chance to display its worst behavior, using parameter values determined in Propositions 1–3.

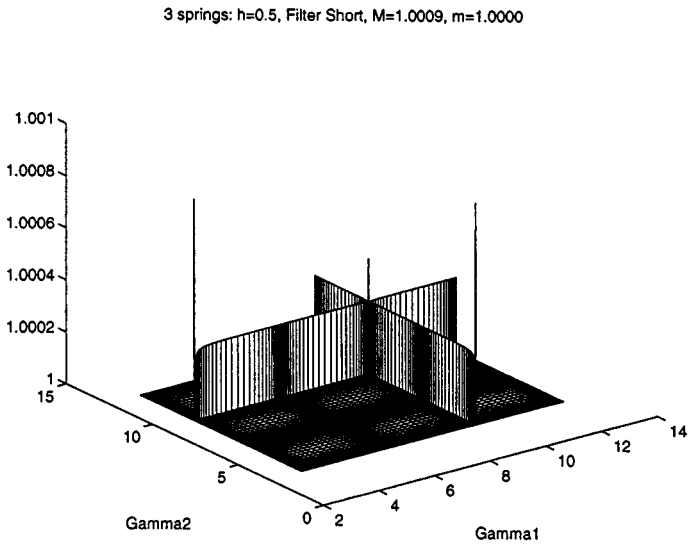

FIG. 5. *Spectral radius of the propagation matrix for ShortAverage method.*

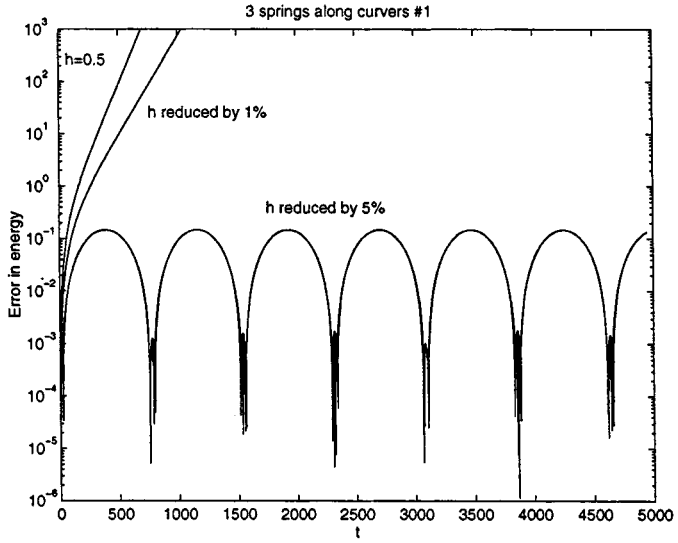

FIG. 6. *Error in energy as a function of time for LongAverage method, linear problem.*

The first set (see Proposition 1) are for

$$\frac{1}{2}\Omega_1 = \frac{\pi}{2} - \frac{1}{4}\left(\frac{1}{2}\right)^2\left(\frac{\pi}{2}\right)^{-3},$$
$$\frac{1}{2}\Omega_2 = \frac{3\pi}{2} - \frac{1}{4}\left(\frac{1}{2}\right)^2\left(\frac{3\pi}{2}\right)^{-3}$$

with initial horizontal momenta of 0.5 and -0.5 for the first and second particles and with zero initial potential energy. Figure 6 shows an exponential growth of the error in energy as a function of time for $h = \frac{1}{2}$. It also shows the error when h is slightly reduced leaving Ω_1 and Ω_2 unchanged. Figure 7 shows what happens when we produce

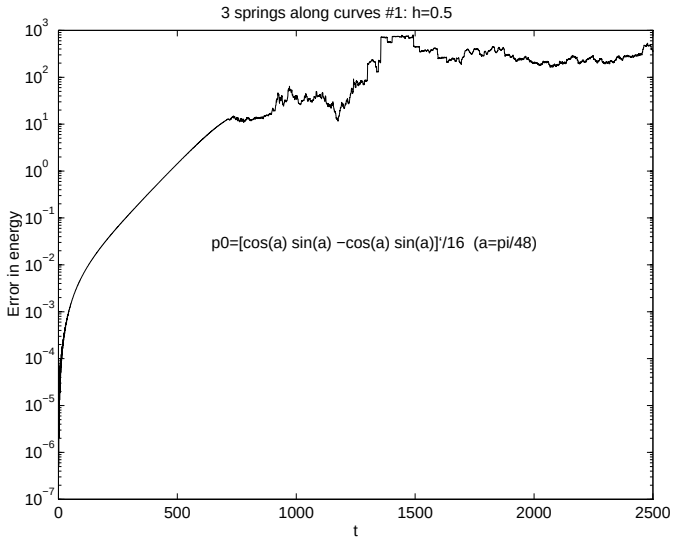

FIG. 7. *Error in energy for LongAverage method for a slightly nonlinear problem.*

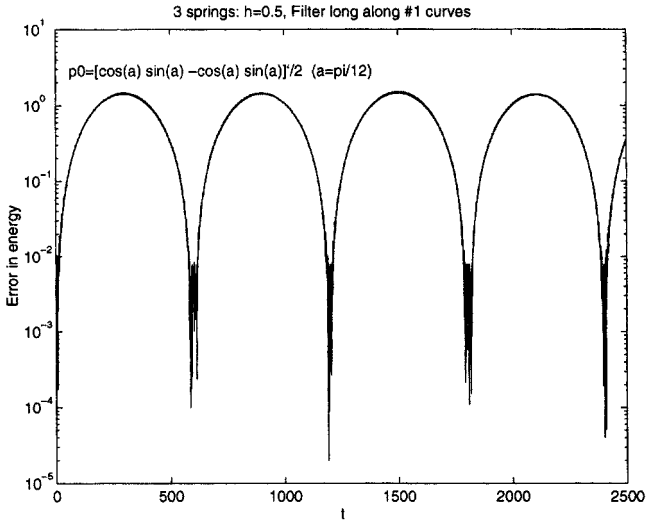

FIG. 8. *Error in energy for LongAverage method for a more nonlinear problem.*

slightly two-dimensional movement of the springs by choosing initial velocities having an angle of $\pi/48$ with the horizontal axis. This yields a nonlinear problem which still exhibits exponential energy growth. Notice that with two-dimensional movement the potential is not defined if a spring length becomes zero, which accounts (we think) for the differences with one-dimensional movement. Figure 8 shows that stability is achieved when a more two-dimensional initial condition is imposed by setting the angle between the horizontal axis and the initial momenta equal to $\pi/12$.

To illustrate instabilities of types 2 and 3, it suffices to use a two-spring problem obtained from our three-spring problem by setting the stiffness of the third spring

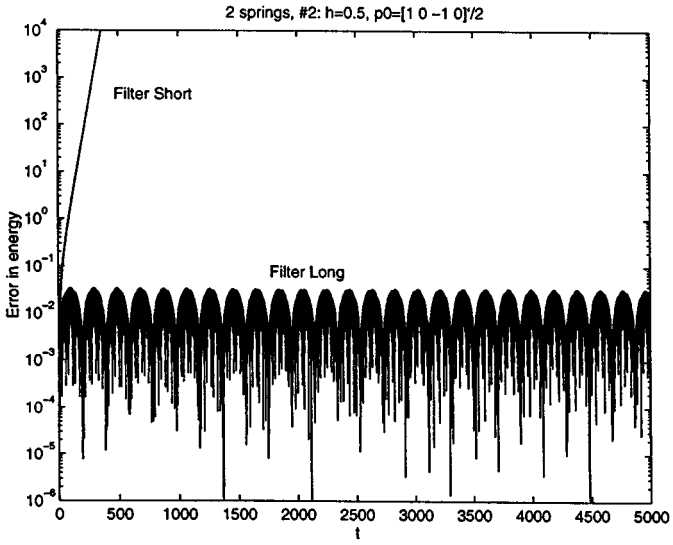

FIG. 9. *Error in energy for LongAverage and ShortAverage methods; testing for instability of type* 2.

to be zero. (The stability analysis of section 7 applies if $\Omega_2 = 0$.) For the following computations we revert to one-dimensional initial conditions.

To get instability of type 2, we choose (see Proposition 2)

$$\frac{1}{2}\Omega_1 = \pi - \left(\frac{1}{2}\right)^2 \pi^{-3}.$$

Figure 9 shows growth of errors in energy for ShortAverage but not LongAverage.

To get instability of type 3, we choose (see Proposition 3)

$$\frac{1}{2}\Omega_1 = 2\pi - \frac{1}{4}\left(\frac{1}{2}\right)^2 (2\pi)^{-1}.$$

Figure 10 shows the growth of errors in energy for Impulse and for ShortAverage but not for LinearAverage nor LongAverage.

4. Accuracy. Section 6 gives a detailed error analysis for the case of linear fast forces, whose results have been summarized in section 1. Here the error analysis is supplemented with numerical experiments on simple but representative nonlinear problems. Again we focus on the Impulse and LongAverage methods and to a lesser extent the ShortAverage method. We use as a test the (nonlinear) two-dimensional spring system described at the beginning of section 3, but we set the stiffness of the third spring equal to zero, which amounts to using a system comprising only the first and second springs.

Figure 11 shows for Impulse the maximum error in position versus Ω_1 for each of step sizes $h = \frac{1}{2}$, $h = \frac{1}{4}$, and $h = \frac{1}{8}$. (As mentioned previously, typical values for the Verlet method in molecular dynamics for h times the *fastest* frequency are from 1/6 to 2/3.) The hard spring has stiffness ranging from 0 to about 1000. Initial conditions are $p = \begin{bmatrix} 1 & 1 & -1 & 1 \end{bmatrix}/(2\sqrt{2})$ and zero potential energy. The time interval is of length 8. For $h = \frac{1}{2}$ the maximum error ≈ 0.07 occurs at $\Omega_1 \approx 4\pi$, while for $h = \frac{1}{4}$ this maximum error ≈ 0.03 occurs at $\Omega_1 \approx 8\pi$. This shows a degraded accuracy when

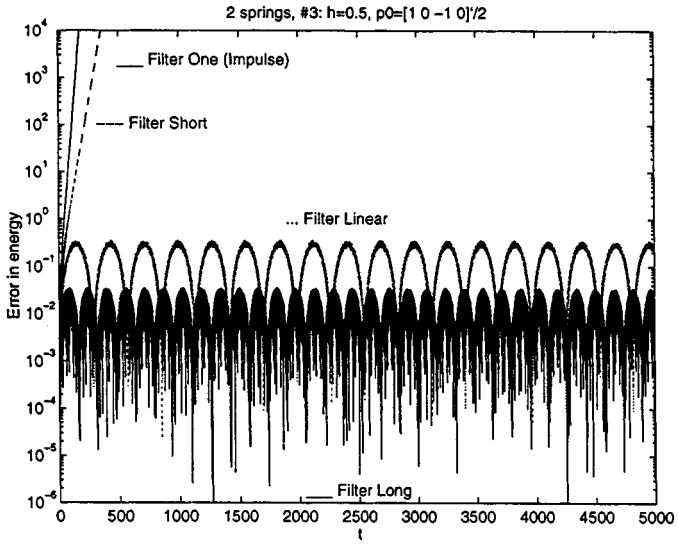

FIG. 10. *Error in energy for LongAverage, LinearAverage, ShortAverage, and Impulse methods; testing for instability of type* 3.

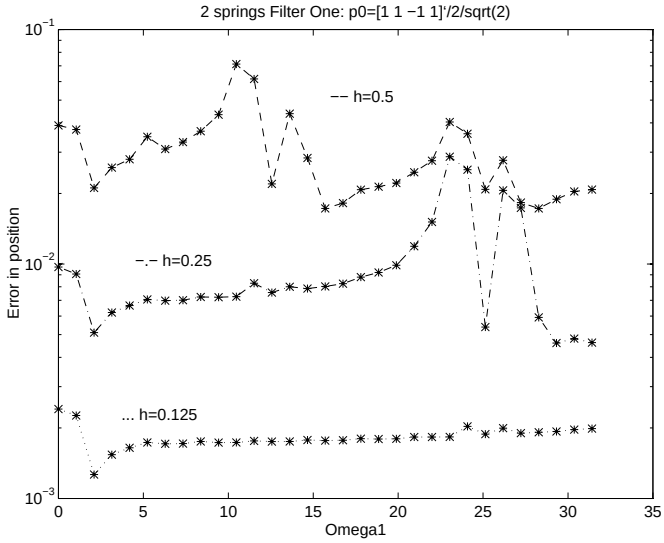

FIG. 11. *Maximum error in position versus* Ω_1 *for Impulse method for each of three different step sizes.*

$h\Omega_1$ is near 2π and, furthermore, that the *maximum* position error (over all possible choices of Ω_1) exhibits an $O(h)$ behavior. This is an *order reduction* of one unit from the expected order 2.

Figure 12 shows the error in position for LongAverage. The time interval is now longer, of length 16. Halving the step size from $\frac{1}{2}$ to $\frac{1}{4}$ reduces the maximum error from 0.2 to 0.05, so that no order reduction is apparent.

A comparison of the accuracy of the three methods, Impulse, ShortAverage, and LongAverage, for $h = \frac{1}{2}$, $h = \frac{1}{4}$ is shown in Figure 13. The time interval is again of length 16.

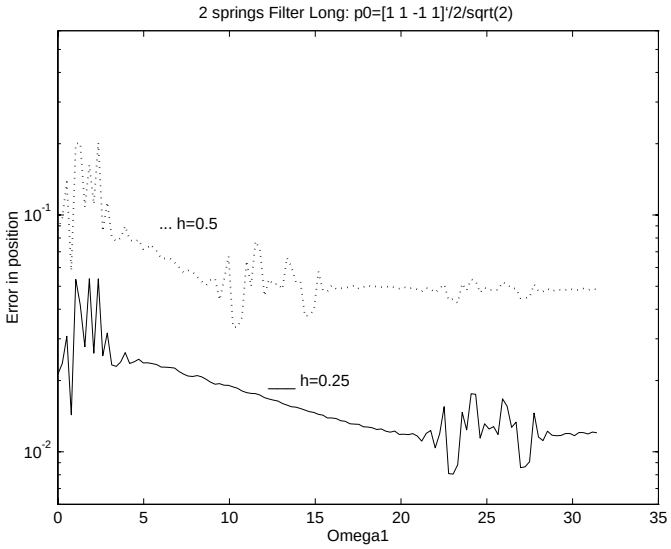

FIG. 12. *Maximum error in position versus* Ω_1 *for LongAverage method for each of two different step sizes.*

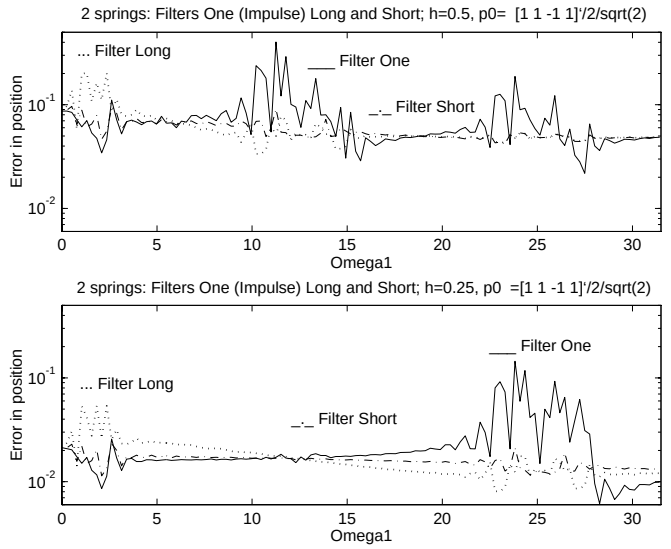

FIG. 13. *Maximum error in position versus* Ω_1 *for LongAverage, ShortAverage, and Impulse methods for each of two different step sizes.*

5. A derivation of the method. We do the derivation in two stages: first, we narrow the set of possibilities by requiring that the scheme be symplectic in the case where the slow force is also a gradient $F = -U_q$; second, we aim for the greatest possible accuracy.

The basic requirement of a long-time-step method is to evaluate F sparingly, so we want to make the best use of each such value. Instead of merely using $p_n^+ = p_n^- + hF(q_n)$, we seek something of the form

$$p_n^+ = p_n^- + h\, \mathit{mollify}(h; q_n, F(\mathcal{A}(h; q_n))),$$

where $\mathcal{A}(h;q_n)$ represents an averaging and where *mollify* produces a softening of the force that more accurately represents the integration of the force into the momentum; cf. (2). These averaging and softening operations are allowed to involve considerable work if the slow forces are expensive to compute. The substep $(p_n^-, q_n) \to (p_n^+, q_n)$ is by itself symplectic if the complete mollified force has a symmetric Jacobian matrix, which is equivalent to it being a gradient. The only obvious way of achieving this, assuming that $F = -U_q$, is by having

$$p_n^+ = p_n^- - h(U(\mathcal{A}(h;q_n)))_q.$$

If we generalize this to nonconservative forces, we have the algorithm that follows: We begin a step with q_n, p_n, and $g_n = F(\mathcal{A}(h;q_n))$ given to us. Then we compute

$$\begin{aligned} p_n^+ &= p_n + \frac{h}{2}\mathcal{A}_q(h;q_n)^{\mathrm{T}} g_n, \\ \begin{bmatrix} q_{n+1} \\ p_{n+1}^- \end{bmatrix} &= h\text{-flow of reduced problem applied to } \begin{bmatrix} q_n \\ p_n^+ \end{bmatrix}, \\ g_{n+1} &= F(\mathcal{A}(h;q_{n+1})), \\ p_{n+1} &= p_{n+1}^- + \frac{h}{2}\mathcal{A}_q(h;q_{n+1})^{\mathrm{T}} g_{n+1}. \end{aligned} \tag{4}$$

The integration of the reduced problem is assumed to be done either exactly or by a symplectic numerical method.

In the belief that the mollification $\mathcal{A}_q(h;q_{n+1})^{\mathrm{T}}$ is more critical than the averaging (for instance averaging does not help if the force is constant), we proceed by determining the best choice for $\mathcal{A}_q(h;q_{n+1})^{\mathrm{T}}$. The aim is to incorporate the available force values g_n not as impulses $h\delta(t-nh)g_n$ but in a more continuous fashion. Ideally we would generate continuous slow forcing functions by interpolation using basis functions $\phi((t-nh)/h)$, meaning that we would integrate the system

$$M\frac{d^2}{dt^2}q = -W_q(q) + \sum_{n=-\infty}^{+\infty} \phi\left(\frac{t-nh}{h}\right) g_n,$$

where $g_n = F(q(nh))$. This is, however, hopelessly implicit and costly in computer time. We derive a much more practical algorithm that involves the integration at every step *of an instance* of the following closely related Hamiltonian system for the extended set of variables (p,q,b,g):

$$\begin{aligned} \frac{d}{dt}p &= -W_q(q) + \phi\left(\frac{t}{h}\right) g, \\ \frac{d}{dt}q &= M^{-1}p, \\ \frac{d}{dt}b &= \phi\left(\frac{t}{h}\right) q, \\ \frac{d}{dt}g &= 0. \end{aligned}$$

Note that in the first equation g contributes to increasing p in a continuous fashion. On the other hand, in any time interval, the increase in b (the variable canonically conjugate to g) is an average of q. The time-dependent Hamiltonian function is

$$H(p,q,b,g,t) = \frac{1}{2}p^{\mathrm{T}}M^{-1}p + W(q) - \phi\left(\frac{t}{h}\right) g^{\mathrm{T}}q.$$

We assume the use of the *same* symplectic *integrator* for this problem[3] as for the reduced problem $(d/dt)p = -W_q(q)$, $(d/dt)q = M^{-1}p$. Let the solution defined by the integrator, with initial values p, q, g, b, be denoted by

$$\begin{bmatrix} P(t;p,q,g) \\ Q(t;p,q,g) \\ b + B(t;q,p,g) \\ g \end{bmatrix}, \quad t = 0, \pm h, \pm 2h, \ldots,$$

whose form implies some very mild assumptions on the integrator. For negative values of t one uses the inverse of the positive flow, which for a reflexive numerical integrator means using the given method with negative h (but for a nonreflexive numerical integrator means using the adjoint method with negative h).

With adequate notation now in place, we present the derivation of the mollified force. Suppose that we have just computed p_n^- and q_n, and we wish to incorporate $\phi(\frac{t-nh}{h})g_n$ into the solution with minimal loss of accuracy. Assume that $\phi(s) = 0$ for $s < -\mu$ and $s > \nu$ for some positive integers μ, ν. The best we can imagine using for p_n^+ is

$$P(-\nu h; Z(\nu h + \mu h; Z(-\mu h; z_n^-, 0), g_n), 0), \tag{5}$$

where $Z = [P^{\mathrm{T}}, Q^{\mathrm{T}}]^{\mathrm{T}}$, $z_n^- = [(p_n^-)^{\mathrm{T}}, q_n^{\mathrm{T}}]^{\mathrm{T}}$, etc. Note that in (5) one integrates forward *with* the force g_n incorporated and backward *without* g_n. The effect of the backward integrations is compensated when flowing forward (rotating) with the reduced flow; hence using (5) for p_n^+ in a kick-rotate-kick algorithm is equivalent to integrating with the term $\phi((t - t_n)/h)g_n$ added to the fast forces.

Therefore, the goal is

$$\begin{aligned} &h\mathcal{A}_q(h;q_n)^{\mathrm{T}}g_n \\ &\approx P(-\nu h; Z(\nu h + \mu h; Z(-\mu h; z_n^-, 0), g_n), 0) - P(-\nu h; Z(\nu h + \mu h; Z(-\mu h; z_n^-, 0), 0), 0), \end{aligned} \tag{6}$$

where the last term is simply a rewriting of p_n^-. We want this to be independent of p_n^-, so we approximate z_n^- by $[q_n^{\mathrm{T}}, 0^{\mathrm{T}}]^{\mathrm{T}}$ in (6). Then to obtain the form of the LHS of (6), we linearize with respect to g_n, obtaining

$$h\mathcal{A}_q(h;q_n)^{\mathrm{T}}g_n = P_z(-\nu h; Z(\nu h), 0) \cdot Z_g(\nu h + \mu h; Z(-\mu h), 0) \cdot g_n,$$

where $Z(t) = Z(t; q_n, 0, 0)$. The possibility of satisfying this equality is justified by the lemma that follows, which directs us to choose

$$\mathcal{A}(h;q) = \frac{1}{h}\left(B(\nu h; q, 0, 0) - B(-\mu h; q, 0, 0)\right).$$

LEMMA 1. *We have that*

$$P_z(-\nu h; Z(\nu h), 0) \cdot Z_g(\nu h + \mu h; Z(-\mu h), 0) = B_q(\nu h; q, 0, 0)^{\mathrm{T}} - B_q(-\mu h; q, 0, 0)^{\mathrm{T}}.$$

Proof. Because $Z(\nu h + \mu h; Z(-\mu h; q, 0, g), g) = Z(\nu h; q, 0, g)$,

$$Z_z(\nu h + \mu h; Z(-\mu h), 0) \cdot Z_g(-\mu h) + Z_g(\nu h + \mu h; Z(-\mu h), 0) = Z_g(\nu h). \tag{7}$$

[3] A numerical integrator defined for autonomous Hamiltonian systems is extended in a natural way to nonautonomous Hamiltonian systems by applying it to an extended Hamiltonian system in which t is treated as an additional position variable; cf. [13].

Similarly $Z(\nu h + \mu h; Z(-\mu h; z, 0), 0) = Z(\nu h; z, 0)$ implies

$$Z_z(\nu h + \mu h; Z(-\mu h), 0) \cdot Z_z(-\mu h) = Z_z(\nu h), \tag{8}$$

and $Z(-\nu h; Z(\nu h; z, 0), 0) = z$ implies

$$Z_z(-\nu h; Z(\nu h), 0) \cdot Z_z(\nu h) = I. \tag{9}$$

Putting together (7), (8), and (9), we have

$$\begin{aligned} &P_z(-\nu h; Z(\nu h), 0) \cdot Z_g(\nu h + \mu h; Z(-\mu h), 0) \\ &\quad = \begin{bmatrix} I & 0 \end{bmatrix} \left(Z_z(\nu h)^{-1} Z_g(\nu h) - Z_z(-\mu h)^{-1} Z_g(-\mu h) \right). \end{aligned}$$

Because the flow

$$\begin{bmatrix} Z(t; z, g) \\ b + B(t; z, g) \\ g \end{bmatrix}$$

is symplectic,

$$\begin{bmatrix} Z_z & 0 & Z_g \\ B_z & I & B_g \\ 0 & 0 & I \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} J^{-1} & 0 & 0 \\ 0 & 0 & -I \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} Z_z & 0 & Z_g \\ B_z & I & B_g \\ 0 & 0 & I \end{bmatrix} = \begin{bmatrix} J^{-1} & 0 & 0 \\ 0 & 0 & -I \\ 0 & I & 0 \end{bmatrix}.$$

From the (1, 1)-block entry of this equation, we get

$$Z_z^{\mathrm{T}} J^{-1} Z_z = J^{-1},$$

and from the (1, 3)-block entry we get

$$Z_z^{\mathrm{T}} J^{-1} Z_g - B_z^{\mathrm{T}} = 0,$$

whence

$$J^{-1} Z_z^{-1} Z_g = B_z^{\mathrm{T}}$$

and

$$\begin{bmatrix} I & 0 \end{bmatrix} Z_z^{-1} Z_g = B_q^{\mathrm{T}}. \qquad \square$$

We have, in section 2, given the details of the mollified impulse method for the leapfrog integrator. In the remainder of this section we discuss two other cases:

1. analytical integration, for the special case of linear fast forces, and
2. integration by Rowlands' method [11].

The first of these provided the insights that originally motivated this method and is of practical interest if spectral methods can be used. The second of these is not only a particularly efficient numerical integrator, but it also illustrates that it is not always obvious how to do the averaging $\mathcal{A}$ without the systematic construction that has just been given.

For analytical integration we have, of course,

$$\mathcal{A}(h; q) = \frac{1}{h} \int_{-\mu h}^{\nu h} \phi\left(\frac{t}{h}\right) Q(t; q, 0) dt.$$

In the case of linear fast forces, $W(q)$ is quadratic $\frac{1}{2}q^TAq + Bq + c$. If we assume also that $W(q)$ has zero as its minimum value, then $A = \Omega^2$ with Ω^2 symmetric semipositive definite, and it can be shown that there exists a vector q^* such that

$$W(q) = \frac{1}{2}(q - q^*)^{\mathrm{T}}\Omega^2(q - q^*).$$

Note that q^* is not normally given explicitly in the natural formulation of the problem and is generally underdetermined. The equations of motion are therefore

$$\frac{d}{dt}p = -\Omega^2(q - q^*) + F(q), \quad \frac{d}{dt}q = M^{-1}p.$$

For theoretical purposes (at least) it is much more convenient if we work with the symplectically transformed system

$$\frac{d}{dt}\bar{p} = -\bar{\Omega}^2\bar{q} + \bar{F}(\bar{q}), \quad \frac{d}{dt}\bar{q} = \bar{p}, \tag{10}$$

where $\bar{p} = M^{-1/2}p$, $\bar{q} = M^{1/2}(q - q^*)$, $\bar{\Omega}^2 = M^{-1/2}\Omega^2M^{-1/2}$, and $\bar{F}(\bar{q}) = M^{-1/2}F(q^* + M^{-1/2}\bar{q})$. In further discussion of the linear case, we work with the transformed system and omit the bars from our notation. We also assume when convenient that the positions and momenta have been orthogonally (and hence symplectically) transformed so that Ω^2 is a nonnegative diagonal matrix. For the transformed system (10) (with the bars now omitted), the flow is governed by

$$\exp\left(t\begin{bmatrix} 0 & -\Omega^2 \\ I & 0 \end{bmatrix}\right) = \begin{bmatrix} \cos t\Omega & -\Omega\sin t\Omega \\ \Omega^{-1}\sin t\Omega & \cos t\Omega \end{bmatrix}.$$

This formula can be used for the full system

$$\frac{d}{dt}p = -\Omega^2 q + F(q), \quad \frac{d}{dt}q = p \tag{11}$$

to express its solution as

$$\begin{aligned}\begin{bmatrix} p(t) \\ q(t) \end{bmatrix} &= \begin{bmatrix} \cos(t-nh)\Omega & -\Omega\sin(t-nh)\Omega \\ \Omega^{-1}\sin(t-nh)\Omega & \cos(t-nh)\Omega \end{bmatrix}\begin{bmatrix} p(nh) \\ q(nh) \end{bmatrix} \\ &\quad + \int_{nh}^{t}\begin{bmatrix} \cos(t-\tau)\Omega \\ \Omega^{-1}\sin(t-\tau)\Omega \end{bmatrix}F(q(\tau))d\tau.\end{aligned} \tag{12}$$

From this we see that

$$Q(t; q, 0) = \cos t\Omega \cdot q$$

and

$$\mathcal{A}(h; q) = \Phi q,$$

where Φ is the matrix

$$\Phi = \frac{1}{h}\int_{-\infty}^{+\infty}\phi\left(\frac{t}{h}\right)\cos t\Omega\, dt = \int_{-\infty}^{+\infty}\phi(s)\cos sh\Omega\, ds.$$

For the methods studied, the "filters" Φ are as follows:

$$\begin{aligned}
&\text{Impulse} && \Phi = I,\\
&\text{ShortAverage} && \Phi = \frac{\sin\frac{h}{2}\Omega}{\frac{h}{2}\Omega},\\
&\text{LongAverage} && \Phi = \frac{\sin h\Omega}{h\Omega} = \frac{\sin\frac{h}{2}\Omega}{\frac{h}{2}\Omega}\cos\frac{h}{2}\Omega,\\
&\text{LinearAverage} && \Phi = \left(\frac{\sin\frac{h}{2}\Omega}{\frac{h}{2}\Omega}\right)^2.
\end{aligned}$$

We conclude this section with a brief look at the application of Rowlands' method [11, 10] to

$$\frac{d}{dt}\begin{bmatrix} p \\ b \end{bmatrix} = \begin{bmatrix} -W_q(q) + \phi(\frac{t}{h})g \\ \phi(\frac{t}{h})q \end{bmatrix}, \quad \frac{d}{dt}\begin{bmatrix} q \\ g \end{bmatrix} = \begin{bmatrix} M^{-1}p \\ 0 \end{bmatrix}.$$

The momentum update has the form

$$\begin{bmatrix} p \\ b \end{bmatrix} := \begin{bmatrix} p \\ b \end{bmatrix} + \frac{\delta t}{2}\left(I - \frac{\delta t^2}{12}\begin{bmatrix} W_{qq}(q) & -\phi(\frac{t}{h}) \\ -\phi(\frac{t}{h}) & 0 \end{bmatrix}\begin{bmatrix} M^{-1} & 0 \\ 0 & 0 \end{bmatrix}\right)\begin{bmatrix} -W_q(q) + \phi(\frac{t}{h})g \\ \phi(\frac{t}{h})q \end{bmatrix}.$$

Taking into account the initialization $g := 0$, the equation for b becomes

$$b := b + \frac{\delta t}{2}\phi\left(\frac{t}{h}\right)\left(q + \frac{\delta t^2}{12}M^{-1}W_q(q)\right),$$

a rather unexpected averaging of the q values!

6. Error bounds and convergence. This section demonstrates second-order accuracy for positions and first-order accuracy for momenta—uniform with respect to the fast force—for the mollified impulse method applied to the case, equation (11), of linear fast forces. Also given are examples of the order reduction of the impulse method and of why we need not only to mollify the slow force but also to average the position values at which this force is evaluated.

The error analysis assumes that in the Euclidean norm $F(q)$ has a Lipschitz constant L_1, that $F(q)$ is bounded by L_0, and that $F_q(q)$ has a Lipschitz constant L_2. We assume (3), that $\phi(s)$ vanishes for $|s| > 1$ and that $\phi(-s) = \phi(s)$. The analysis also requires bounds on $\phi(s)$. For simplicity we assume

$$\int_{-1}^{1} |s^j\phi(s)|ds \le \frac{1}{j+1}, \tag{13}$$

a bound which is satisfied by all three basis functions that have been explicitly mentioned. Larger bounds on $\phi(s)$ lead merely to larger bounds on the error. The bound we obtain is in terms of the "reduced energy"

$$\hat{H} = \max_{0\le t\le nh}\left(\frac{1}{2}p(t)^{\mathrm{T}}p(t) + \frac{1}{2}q(t)^{\mathrm{T}}\Omega^2 q(t)\right).$$

THEOREM 1. *Let $t = nh$. Under the preceding hypotheses, the global error of the mollified methods satisfies*

$$\|q_n - q(nh)\| \le h^2\cosh(tL_1)\cdot\left(\frac{1}{2}L_0 + \frac{13}{12}L_1\hat{H}^{1/2}t + \frac{1}{2}t^2\left(\frac{1}{6}L_1L_0 + \frac{4}{3}L_2\hat{H}\right)\right) \tag{14}$$

and

$$\begin{aligned}\|p_n - p(nh)\| \le{}& h\left(\frac{1}{2}L_0 + (2t+h)L_1\hat{H}^{1/2}\right)\\ &+ h^2 L_1 t\cosh(tL_1)\cdot\left(\frac{1}{2}L_0 + \frac{13}{12}L_1\hat{H}^{1/2}t + \frac{1}{2}t^2\left(\frac{1}{6}L_1L_0 + \frac{4}{3}L_2\hat{H}\right)\right)\\ &+ h^2 t\left(\frac{1}{6}L_1L_0 + \frac{4}{3}L_2\hat{H}\right).\end{aligned} \tag{15}$$

Before we prove the theorem, we discuss some counterexamples.

First we note that the example around equation (2) shows that for the impulse method it is not possible to bound $\|p_n - p(t_n)\|$ by Ch with the constant C depending on $\hat{H}$, L_0, L_1, L_2, and t as in (15). Hence the impulse method undergoes an order reduction in p from 2 to 0.

Next we consider the three-spring system of section 3 with one-dimensional initial conditions and $\Omega_2 = 0$. We denote by P and p the (horizontal) momenta of the first and second masses and by $1+Q$ and $2+q$ the corresponding abscissae. As $\Omega_1 \to \infty$, the frequencies in the system are

$$\Omega_+ = \Omega_1 + \frac{1}{4\Omega_1} + O\left(\frac{1}{\Omega_1^3}\right), \quad \Omega_- = \frac{\sqrt{2}}{2} + O\left(\frac{1}{\Omega_1^2}\right). \tag{16}$$

As h is reduced in the numerical methods, we increase Ω_1 so as to keep

$$h\Omega_1 = 2\pi. \tag{17}$$

Assume that we integrate either with the impulse method or with a method that is mollified with the Short, Long, or Linear filters but where no average is done to evaluate the force (i.e., with a method using the force $\mathcal{A}_q(h;q_n)^T F(q_n)$). In this setting, it is easy to check the following points.

(i) The numerical value of Q_n does not change along the integration, i.e., $Q_n = Q_0$.

(ii) The variables $\chi_n = Q_n - q_n$ and p_n obey the equations

$$\begin{aligned}p_n^+ &= p_n^- - \frac{h}{4}\chi_n,\\ \chi_{n+1} &= \chi_n + hp_n^+,\\ p_{n+1} &= p_n^+ - \frac{h}{4}\chi_{n+1}.\end{aligned}$$

These coincide with the Verlet discretization of the system $\dot{p} = -\frac{1}{2}\chi$, $\dot{\chi} = p$ (a single spring with angular frequency $\sqrt{2}/2$).

We choose the initial condition so that the solution is the eigenmode

$$\begin{aligned}P(t) &= -\sin\Omega_+ t,\\ p(t) &= \frac{1}{2\Omega_+^2 - 1}\sin\Omega_+ t,\\ Q(t) &= \frac{1}{\Omega_+}\cos\Omega_+ t,\\ q(t) &= -\frac{1}{\Omega_+(2\Omega_+^2 - 1)}\cos\Omega_+ t.\end{aligned}$$

Note that the energy in this solution remains bounded as $\Omega_1 \to \infty$. In view of (17) and (16), in the true solution $q(t) = O(h^3)$. However for the numerical solution, (i)–(ii) above and the convergence of the Verlet method imply

$$\begin{aligned} q_n &= Q_n + \chi_n \\ &= Q(0) + \chi(t_n) + O(h^2) \\ &= Q(0) + \chi(0)\cos\frac{\sqrt{2}}{2}t_n + O(h^2) \\ &= \frac{1}{\Omega_+}\left(1 - \cos\frac{\sqrt{2}}{2}t_n\right) + O(h^2) = O(h). \end{aligned}$$

Hence q_n is an error by an $O(h)$ amount. We conclude that the impulse method undergoes an order reduction in q from 2 to 1. Furthermore, mollification without averaging inside F is not sufficient to ensure an $O(h^2)$ bound like (14).

A similar counterexample, involving the same spring system, can be constructed to show that the $O(h)$ bound in (15) cannot be improved to become $O(h^2)$.

We now prove Theorem 1. We first need a discrete Gronwall lemma appropriate for special second-order ordinary differential equations.

LEMMA 2. *Let*

$$E_n \le D_n + \eta \sum_{j=1}^{n-1} (n-j)E_j, \quad n = 1, 2, \ldots. \tag{18}$$

Then

$$E_n \le D_n + \eta \sum_{j=1}^{n-1} \frac{\rho^{n-j} - \rho^{j-n}}{\rho - \rho^{-1}} D_j, \tag{19}$$

where

$$\rho = 1 + \frac{1}{2}\eta + \sqrt{\eta + \frac{1}{4}\eta^2}.$$

Moreover,

$$E_n \le \cosh(n\sqrt{\eta}) \cdot \max_{1 \le j \le n} D_j.$$

Proof. The inequality (18) is majorized by the solution of

$$\bar{E}_n = D_n + \eta \sum_{j=1}^{n-1} (n-j)\bar{E}_j.$$

A lengthy calculation shows the solution to be the RHS of inequality (19). It can be checked fairly easily by induction using the fact that $(\rho - 1)^2 = \eta\rho$. We have

$$\rho \le 1 + \eta^{1/2} + \frac{1}{2}\eta + \frac{1}{8}\eta^{3/2} \le \exp(\eta^{1/2}).$$

Let $D = \max_k D_k$. Then by computing the sum on the RHS of (19) and using $(\rho - 1)^2 = \eta\rho$, we get

$$\begin{aligned} E_n &\le \frac{\rho^n + \rho^{1-n}}{\rho + 1} D \le \frac{\rho^{n-1/2} + \rho^{1/2-n}}{2} D \\ &= \cosh\left(\left(n - \frac{1}{2}\right)\log\rho\right) \cdot D \le \cosh\left(\left(n - \frac{1}{2}\right)\eta^{1/2}\right) \cdot D. \end{aligned}$$

∎

Now we are equipped to obtain a bound on the global error in terms of quadrature error.

LEMMA 3. *The global error satisfies*

$$\|q_n - q(nh)\| \leq \cosh(L_1 nh) \cdot \max_{1\leq j\leq n} \|\sigma_{q,j}\|$$

and

$$\|p_n - p(nh)\| \leq L_1 nh \cosh(L_1 nh) \cdot \max_{1\leq j\leq n} \|\sigma_{q,j}\| + \|\sigma_{p,n}\|$$

where $\sigma_{p,n}$, $\sigma_{q,n}$ are the quadrature errors

$$\begin{bmatrix} \sigma_{p,n} \\ \sigma_{q,n} \end{bmatrix} = h\sum_{j=0}^{n} 1_j R^{n-j} \begin{bmatrix} \Phi \\ 0 \end{bmatrix} F(\Phi q(jh)) - \int_0^{nh} \begin{bmatrix} \cos(nh-t)\Omega \\ \Omega^{-1}\sin(nh-t)\Omega \end{bmatrix} F(q(t))dt$$

with

$$R = \begin{bmatrix} \cos h\Omega & -\Omega \sin h\Omega \\ \Omega^{-1}\sin h\Omega & \cos h\Omega \end{bmatrix}$$

and $1_j = 1$ except that $1_0 = 1_n = \frac{1}{2}$.

Proof. From (12) the analytical solution satisfies

$$\begin{bmatrix} p(nh) \\ q(nh) \end{bmatrix} = R \begin{bmatrix} p((n-1)h) \\ q((n-1)h) \end{bmatrix} + \int_{(n-1)h}^{nh} \begin{bmatrix} \cos(nh-t)\Omega \\ \Omega^{-1}\sin(nh-t)\Omega \end{bmatrix} F(q(t))dt.$$

The numerical solution defined by (4) satisfies

$$\begin{bmatrix} p_n \\ q_n \end{bmatrix} = R \begin{bmatrix} p_{n-1} \\ q_{n-1} \end{bmatrix} + R \begin{bmatrix} \frac{h}{2}\Phi F(\Phi q_{n-1}) \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{h}{2}\Phi F(\Phi q_n) \\ 0 \end{bmatrix}.$$

Letting $\delta_n = p_n - p(nh)$, $\varepsilon_n = q_n - q(nh)$, and $\Delta_n = F(\Phi q_n) - F(\Phi q(nh))$, we have

$$\begin{bmatrix} \delta_n \\ \varepsilon_n \end{bmatrix} = R \begin{bmatrix} \delta_{n-1} \\ \varepsilon_{n-1} \end{bmatrix} + \frac{h}{2} R \begin{bmatrix} \Phi \\ 0 \end{bmatrix} \Delta_{n-1} + \frac{h}{2} \begin{bmatrix} \Phi \\ 0 \end{bmatrix} \Delta_n + \tau_n \tag{20}$$

where

$$\begin{aligned} \tau_n = {} & \frac{h}{2} R \begin{bmatrix} \Phi \\ 0 \end{bmatrix} F(\Phi q((n-1)h)) + \frac{h}{2} \begin{bmatrix} \Phi \\ 0 \end{bmatrix} F(\Phi q(nh)) \\ & - \int_{nh-h}^{nh} \begin{bmatrix} \cos(nh-t)\Omega \\ \Omega^{-1}\sin(nh-t)\Omega \end{bmatrix} F(q(t))dt. \end{aligned}$$

Summing (20) yields

$$\begin{bmatrix} \delta_n \\ \varepsilon_n \end{bmatrix} = h\sum_{j=0}^{n} 1_j R^{n-j} \begin{bmatrix} \Phi \\ 0 \end{bmatrix} \Delta_j + \sigma_n,$$

where

$$\sigma_n = \begin{bmatrix} \sigma_{p,n} \\ \sigma_{q,n} \end{bmatrix} = \sum_{j=1}^{n} R^{n-j}\tau_j.$$

The equation for ε_n, $n \geq 1$, is

$$\varepsilon_n = h \sum_{j=1}^{n-1} \Omega^{-1} \sin(n-j)h\Omega \cdot \Phi\Delta_j + \sigma_{q,n},$$

so

$$\|\varepsilon_n\| \leq \|\sigma_{q,n}\| + h^2 L_1 \sum_{j=1}^{n-1} (n-j)\|\varepsilon_j\|,$$

where we have used (13) to conclude that $\|\Phi\| \leq 1$. Applying Lemma 2 leads to the upper bound given for $\|q_n - q(nh)\|$. The equation for δ_n, $n \geq 1$, is

$$\delta_n = h \sum_{j=1}^{n} 1_j \cos(n-j)h\Omega \cdot \Phi\Delta_j + \sigma_{p,n},$$

so

$$\|\delta_n\| \leq hL_1 \sum_{j=1}^{n} \|\varepsilon_j\| + \|\sigma_{p,n}\|. \qquad \square$$

Proof of Theorem 1. Letting $\sigma_n = \left[\sigma_{p,n}^{\mathrm{T}}, \sigma_{q,n}^{\mathrm{T}}\right]^{\mathrm{T}}$, we can write $\sigma_n = \sigma_n^1 - \sigma_n^2$, where

$$\sigma_n^1 = h \sum_{j=0}^{n} 1_j R^{n-j} \begin{bmatrix} \Phi \\ 0 \end{bmatrix} F_j$$

with $F_j = F(\Phi q(jh))$ and

$$\sigma_n^2 = \int_0^{nh} \begin{bmatrix} \cos(nh-t)\Omega \\ \Omega^{-1}\sin(nh-t)\Omega \end{bmatrix} F(q(t))dt.$$

Because of our assumption that $\phi(s)$ vanishes for $|s| > 1$ and that $\sum_{j=-\infty}^{+\infty} \phi(s-j) \equiv 1$, we can write

$$\begin{aligned}
\sigma_n^2 &= \int_0^{nh} \begin{bmatrix} \cos(nh-t)\Omega \\ \Omega^{-1}\sin(nh-t)\Omega \end{bmatrix} \sum_{j=0}^{n} \phi\left(\frac{t-jh}{h}\right) F(q(t))dt \\
&= \sum_{j=0}^{n} R^{n-j} \int_0^{nh} \begin{bmatrix} \cos(jh-t)\Omega \\ \Omega^{-1}\sin(jh-t)\Omega \end{bmatrix} \phi\left(\frac{t-jh}{h}\right) F(q(t))dt \\
&= \sum_{j=0}^{n} R^{n-j} \int_{-jh}^{(n-j)h} \begin{bmatrix} \cos t\Omega \\ -\Omega^{-1}\sin t\Omega \end{bmatrix} \phi\left(\frac{t}{h}\right) F(q(jh+t))dt.
\end{aligned}$$

By comparison

$$\begin{aligned}
\sigma_n^1 = &\sum_{j=0}^{n} R^{n-j} \int_{-jh}^{(n-j)h} \begin{bmatrix} \cos t\Omega \\ -\Omega^{-1}\sin t\Omega \end{bmatrix} \phi\left(\frac{t}{h}\right) F_j dt \\
&+ R^n \int_{-h}^{0} \begin{bmatrix} 0 \\ -\Omega^{-1}\sin t\Omega \end{bmatrix} \phi\left(\frac{t}{h}\right) F_0 dt + \int_0^h \begin{bmatrix} 0 \\ -\Omega^{-1}\sin t\Omega \end{bmatrix} \phi\left(\frac{t}{h}\right) F_n dt,
\end{aligned}$$

where we have used the fact that $\int_{-h}^{h} \sin t\Omega \cdot \phi(t/h)dt = 0$. Hence

$$\sigma_n = \sigma_n^4 + \sigma_n^3 \tag{21}$$

where

$$\sigma_n^4 = \sum_{j=0}^{n} R^{n-j} \int_{-jh}^{(n-j)h} \begin{bmatrix} \cos t\Omega \\ -\Omega^{-1}\sin t\Omega \end{bmatrix} \phi\left(\frac{t}{h}\right)(F_j - F(q(jh+t)))\,dt$$

and

$$\begin{aligned} \begin{bmatrix} \|\sigma_{p,n}^3\| \\ \|\sigma_{q,n}^3\| \end{bmatrix} &\le \int_{-h}^{0} \begin{bmatrix} 1 \\ |t| \end{bmatrix} \left|\phi\left(\frac{t}{h}\right)\right| dt \cdot \|F_0\| + \int_0^h \begin{bmatrix} 0 \\ |t| \end{bmatrix} \left|\phi\left(\frac{t}{h}\right)\right| dt \cdot \|F_n\| \\ &\le \begin{bmatrix} 1 \\ h \end{bmatrix} \frac{h}{2} L_0. \end{aligned} \tag{22}$$

It remains to bound σ_n^4, so we write

$$\sigma_n^4 = \sum_{j=0}^{n} R^{n-j} \left(\tau_j^5 - \begin{bmatrix} \tau_{p,j}^6 \\ 0 \end{bmatrix}\right), \tag{23}$$

where

$$\tau_{p,j}^6 = \int_{-jh}^{(n-j)h} \phi\left(\frac{t}{h}\right)(F(q(jh+t)) - F_j)\,dt$$

and

$$\tau_j^5 = \int_{-jh}^{(n-j)h} \begin{bmatrix} \sin\frac{t}{2}\Omega \\ \Omega^{-1}\cos\frac{t}{2}\Omega \end{bmatrix} 2\sin\frac{t}{2}\Omega \cdot \phi\left(\frac{t}{h}\right)(F(q(jh+t) - F_j))\,dt.$$

To bound τ_j^5, we write

$$\begin{aligned} q(jh+t) - \Phi q(jh) &= \frac{1}{h}\int_{-h}^{h} \phi\left(\frac{t}{h}\right)(1-\cos t\Omega)dt \cdot q(jh) + \int_0^t p(jh+\tau)d\tau \\ &= \frac{1}{h}\int_{-h}^{h} t\phi\left(\frac{t}{h}\right)\sin\frac{t}{2}\Omega\frac{\sin\frac{t}{2}\Omega}{\frac{t}{2}\Omega}dt \cdot \Omega q(jh) + \int_0^t p(jh+\tau)d\tau, \end{aligned}$$

from which we get

$$\|q(jh+t) - \Phi q(jh)\| \le \left(\frac{h}{2} + |t|\right)\hat{H}^{1/2}$$

and

$$\begin{aligned} \|R^{n-j}\tau_j^5\| &\le 1_j \int_{-h}^{h} \begin{bmatrix} 2 \\ |t| \end{bmatrix} \left|\phi\left(\frac{t}{h}\right)\right| L_1\left(\frac{h}{2} + |t|\right)\hat{H}^{1/2}dt \\ &\le 1_j \begin{bmatrix} 2 \\ \frac{7}{12}h \end{bmatrix} h^2 L_1 \hat{H}^{1/2}. \end{aligned} \tag{24}$$

It remains to bound $\tau_{p,j}^6$. We have

$$F(q(jh+t)) = F_j + F_{q,j}(q(jh+t) - \Phi q(jh)) + E_j(t),$$

where

$$F_{q,j} = F_q(\Phi q(jh))$$

and

$$\begin{aligned}\|E_j(t)\| &\le \frac{1}{2}L_2\|q(jh+t)-\Phi q(jh)\|^2\\ &\le \frac{1}{2}L_2\left(\frac{h}{2}+|t|\right)^2\hat{H}.\end{aligned}$$

Also, we have that

$$q(jh+t)=\cos t\Omega\cdot q(jh)+\Omega^{-1}\sin t\Omega\cdot p(jh)+\Delta_j(t)$$

where

$$\Delta_j(t)=\int_0^t \Omega^{-1}\sin(t-s)\Omega\cdot F(q(jh+s))ds.$$

This last quantity satisfies

$$\|\Delta_j(t)\|\le\frac{1}{2}t^2L_0.$$

Therefore,

$$\tau^6_{p,j}=\tau^7_{p,j}+\tau^8_{p,j}+\tau^9_{p,j},$$

where

$$\begin{aligned}\tau^7_{p,j}&=F_{q,j}\int_{-jh}^{(n-j)h}\phi\left(\frac{t}{h}\right)(\cos t\Omega-\Phi)dt\cdot q(jh),\\ \tau^8_{p,j}&=F_{q,j}\int_{-jh}^{(n-j)h}\phi\left(\frac{t}{h}\right)\Omega^{-1}\sin t\Omega dt\cdot p(jh),\end{aligned}$$

and

$$\tau^9_{p,j}=\int_{-jh}^{(n-j)h}\phi\left(\frac{t}{h}\right)\left(F_{q,j}\Delta_j(t)+E_j(t)\right)dt.$$

The last of these admits the bound

$$\begin{aligned}\|\tau^9_{p,j}\|&\le h^3\int_{-j}^{n-j}|\phi(s)|\left(\frac{1}{2}s^2L_1L_0+\frac{1}{2}L_2\left(\frac{h}{2}+|s|\right)^2\hat{H}\right)ds\\ &\le h^3 1_j\left(\frac{1}{6}L_1L_0+\frac{4}{3}L_2\hat{H}\right).\end{aligned}$$

For $0<j<n$ we have $\tau^8_{p,j}=0$ because it is the integral from $-h$ to $+h$ of an odd function, and otherwise we have

$$\|\tau^8_{p,j}\|\le\frac{1}{2}h^2L_1\hat{H}^{1/2}.$$

Finally,

$$\int_0^h\phi\left(\frac{t}{h}\right)(\cos t-\Phi)\,dt=\frac{1}{2}\Phi-\frac{1}{2}\Phi=0,$$

and similarly for the integral from $-h$ to 0. This implies $\tau_j^7 = 0$. Combining (21), (22), (23), and (24) yields

$$\begin{aligned}\begin{bmatrix} \|\sigma_{p,n}\| \\ \|\sigma_{q,n}\| \end{bmatrix} &\le \begin{bmatrix} 1 \\ h \end{bmatrix} \frac{h}{2} L_0 + \sum_{j=0}^{n} \left(1_j \begin{bmatrix} 2 \\ \frac{7}{12}h \end{bmatrix} h^2 L_1 \hat{H}^{1/2} + \begin{bmatrix} \|\cos(n-j)h\Omega \cdot \tau_{p,j}^6\| \\ \|\Omega^{-1}\sin(n-j)h\Omega \cdot \tau_{p,j}^6\| \end{bmatrix} \right) \\ &\le \begin{bmatrix} 1 \\ h \end{bmatrix} \frac{h}{2} L_0 + \begin{bmatrix} 2 \\ \frac{7}{12}h \end{bmatrix} thL_1 \hat{H}^{1/2} + \sum_{j=0}^{n} \begin{bmatrix} 1 \\ (n-j)h \end{bmatrix} \|\tau_{p,j}^6\|. \end{aligned} \tag{25}$$

Combining the bounds for $\tau_{p,j}^9$, $\tau_{p,j}^8$, and $\tau_{p,j}^7$ gives

$$\|\tau_{p,j}^6\| \le (1-1_j)h^2 L_1 \hat{H}^{1/2} + h^3 1_j \left(\frac{1}{6} L_1 L_0 + \frac{4}{3} L_2 \hat{H} \right). \tag{26}$$

The theorem follows from (25), (26), and Lemma 3 using the fact that $\sum_{j=0}^{n}(n-j)1_j = \frac{1}{2}n^2$. ∎

7. Stability analysis. Here we provide the details that support the assertions in section 3.

We consider the two-degree-of-freedom linear problem

$$\frac{d}{dt}p = -\Omega^2 q - U_{qq} q, \quad \frac{d}{dt}q = p, \tag{27}$$

where, without loss of generality, $\Omega^2 = \operatorname{diag}(\Omega_1^2, \Omega_2^2)$ and

$$U_{qq} = \begin{bmatrix} \alpha_1 & \beta \\ \beta & \alpha_2 \end{bmatrix}$$

is assumed to be a constant symmetric semipositive definite matrix. To analyze stability, we need the characteristic polynomial of the matrix that propagates the solution and also the errors.

LEMMA 4. *The eigenvalues λ of the error propagation matrix for integrator* (4) *applied to* (27) *satisfy*

$$\lambda^4 - 2S\lambda^3 + (2 + S^2 - D)\lambda^2 - 2S\lambda + 1 = 0, \tag{28}$$

where S and D are given by

$$\begin{aligned} S &= \cos\Gamma_1 - \varepsilon_1 + \cos\Gamma_2 - \varepsilon_2, \\ D &= (\cos\Gamma_1 - \varepsilon_1 - \cos\Gamma_2 + \varepsilon_2)^2 + 4\theta^2 \varepsilon_1 \varepsilon_2 \end{aligned}$$

with

$$\varepsilon_i = \frac{1}{2} h^2 \alpha_i \Phi(\Gamma_i)^2 \Gamma_i^{-1} \sin\Gamma_i,$$

$\Gamma_i = h\Omega_i$, *and* $\theta = \beta/\sqrt{\alpha_1\alpha_2}$.

Proof. One step of the mollified impulse method (4) is given by

$$\begin{bmatrix} p_{n+1} \\ q_{n+1} \end{bmatrix} = A \begin{bmatrix} p_n \\ q_n \end{bmatrix},$$

where

$$A = \begin{bmatrix} I & -\frac{h}{2}\Phi U_{qq}\Phi \\ 0 & I \end{bmatrix} \begin{bmatrix} \cos h\Omega & -\Omega \sin h\Omega \\ \Omega^{-1}\sin h\Omega & \cos h\Omega \end{bmatrix} \begin{bmatrix} I & -\frac{h}{2}\Phi U_{qq}\Phi \\ 0 & I \end{bmatrix}.$$

The error propagation matrix A is similar to

$$A' = \begin{bmatrix} I & -h^2\Phi U_{qq}\Phi \\ 0 & I \end{bmatrix} \begin{bmatrix} \cos h\Omega & -h\Omega \sin h\Omega \\ (h\Omega)^{-1}\sin h\Omega & \cos h\Omega \end{bmatrix}.$$

The lemma follows from a calculation of $\det(\lambda I - A')$ using $\Phi = \mathrm{diag}(\Phi(\Gamma_1), \Phi(\Gamma_2))$, $\cos h\Omega = \mathrm{diag}(\cos\Gamma_1, \cos\Gamma_2)$, and $\sin h\Omega = \mathrm{diag}(\sin\Gamma_1, \sin\Gamma_2)$. ∎

Recall that the filters are

$$\Phi(\Gamma) = \begin{cases} 1, & \text{Impulse}, \\ \frac{\sin(\Gamma/2)}{\Gamma/2}, & \text{ShortAverage}, \\ \frac{\sin\Gamma}{\Gamma}, & \text{LongAverage}. \end{cases}$$

Next we obtain necessary conditions for stability.

LEMMA 5. *The roots of* (28) *are less than or equal to one in modulus if and only if*

1. $D \geq 0$,
2. $S \geq -2 + \sqrt{D}$, *and*
3. $S \leq 2 - \sqrt{D}$.

Conditions 2 *and* 3 *can be combined as* $|S| \leq 2 - \sqrt{D}$.

Proof. Assume that the roots of (28) are less than or equal to one in modulus. Because their product is one, they must all be equal to one in modulus. Hence it is necessary that the quartic of (28) be factorizable as

$$(\lambda^2 + a\lambda + 1)(\lambda^2 + b\lambda + 1) = \lambda^4 + (a+b)\lambda^3 + (ab+2)\lambda^2 + (a+b)\lambda + 1,$$

where $-2 \leq a, b \leq 2$. Comparing this quartic with that of (28), we note that a and b must be the roots of $x^2 + 2Sx + S^2 - D = 0$, so

$$a, b = -S \pm \sqrt{D}.$$

Hence it is necessary that $D \geq 0$. Additionally, it is necessary that $-2 \leq -S \pm \sqrt{D} \leq 2$. This proves necessity of conditions 1–3; sufficiency is proved by a straightforward reversal of the preceding arguments. ∎

For $U_{qq} = 0$ the roots of (28) are all equal to one in modulus. What happens as U_{qq} grows is that roots coalesce and then bifurcate off the unit circle. Where this happens determines which of conditions 1–3 is violated:

type 1 instability occurs if $D < 0$ and is associated with the coalescing of imaginary eigenvalues at an imaginary point. The boundary of a type-1 instability region for fixed h is contained in the set of (Γ_1, Γ_2) which satisfy

$$D = 0. \tag{29}$$

type 2 instability occurs in regions where $D \geq 0$ if $S - \sqrt{D} < -2$. It is associated with the coalescing of imaginary eigenvalues at -1. Its boundary is contained in the set of (Γ_1, Γ_2) which satisfies

$$(1 + \cos\Gamma_1 - \varepsilon_1)(1 + \cos\Gamma_2 - \varepsilon_2) = \theta^2\varepsilon_1\varepsilon_2. \tag{30}$$

type 3 instability occurs in regions where $D \geq 0$ if $S + \sqrt{D} > 2$. It is associated with the coalescing of imaginary eigenvalues at $+1$. Its boundary is contained in the set of (Γ_1, Γ_2) which satisfy

$$(1 - \cos\Gamma_1 + \varepsilon_1)(1 - \cos\Gamma_2 + \varepsilon_2) = \theta^2\varepsilon_1\varepsilon_2. \tag{31}$$

The proposition that follows is the result of an attempt to find the worst possible case of type-1 instability. In the interests of brevity we do not prove that it is the worst case. First a lemma is needed.

LEMMA 6. *If $D < 0$, then*

$$a = \frac{1}{4}\left(\sqrt{(2+S)^2 - D} + \sqrt{(2-S)^2 - D}\right) > 1,$$

and the largest root of the LHS of (28) *has modulus*

$$a + \sqrt{a^2 - 1}.$$

If $|S| < 2$, then the modulus is

$$1 + \sqrt{\frac{|D|}{4 - S^2}} + O\left(\frac{|D|}{2 - |S|}\right).$$

Proof. We begin by noting that

$$a > \frac{1}{4}(|2+S| + |2-S|) \geq 1.$$

With $\rho = a + \sqrt{a^2 - 1}$ and

$$\sigma = \frac{1}{2}\left(\sqrt{(2+S)^2 - D} - \sqrt{(2-S)^2 - D}\right),$$

we can by direct calculation show that the LHS of (28) factors as

$$(\lambda^2 - \sigma\rho\lambda + \rho^2)(\lambda^2 - \sigma\rho^{-1}\lambda + \rho^{-2}).$$

(It helps to use the fact that $\rho + \rho^{-1} = 2a$.) Because $\rho > 1$, the root λ of maximum modulus is one of

$$\lambda = \rho\left(\frac{\sigma}{2} \pm \sqrt{\left(\frac{\sigma}{2}\right)^2 - 1}\right).$$

Now

$$|\sigma| = \frac{4|S|}{4a} < \frac{2|S+2| + 2|S-2|}{|2+S| + |2-S|} = 2,$$

so λ above has modulus ρ, which proves the first part of the lemma. If $|S| < 2$, then

$$a = 1 - \frac{D}{2(4 - S^2)} + O\left(\frac{D^2}{(2 - |S|)^3}\right),$$

from which the second part follows. ∎

PROPOSITION 1. *Let $\Gamma > 0$ and $\bar{\Gamma} > 0$ be such that $\Gamma + \bar{\Gamma}$ is an integer multiple of 2π. Then let*

$$\Gamma_1 = \Gamma - \frac{1}{2}h^2\alpha_1\Phi(\Gamma)^2\Gamma^{-1},$$
$$\Gamma_2 = \bar{\Gamma} - \frac{1}{2}h^2\alpha_2\Phi(\bar{\Gamma})^2\bar{\Gamma}^{-1}.$$

If Γ is not an integer multiple of π, then the spectral radius of the error propagation matrix is

$$1 + \frac{1}{2}h^2|\beta|\sqrt{\Phi(\Gamma)^2\Gamma^{-1}\Phi(\bar{\Gamma})^2\bar{\Gamma}^{-1}} + O(h^4).$$

Proof. Letting $\Delta = \frac{1}{2}h^2\alpha_1\Phi(\Gamma)^2\Gamma^{-1}$, we have

$$\begin{aligned}
\Gamma_1 &= \Gamma - \Delta, \\
\varepsilon_1 &= \frac{1}{2}h^2\alpha_1\Phi(\Gamma)^2\Gamma^{-1}\sin\Gamma + O(h^4) \\
&= \Delta\sin\Gamma + O(h^4), \\
\cos\Gamma_1 &= \cos\Gamma\cos\Delta + \sin\Gamma\sin\Delta \\
&= \cos\Gamma + \Delta\sin\Gamma + O(h^4),
\end{aligned}$$

so

$$\cos\Gamma_1 - \varepsilon_1 = \cos\Gamma + O(h^4).$$

Similarly

$$\varepsilon_2 = \frac{1}{2}h^2\alpha_2\Phi(\bar{\Gamma})^2\bar{\Gamma}^{-1}\sin\bar{\Gamma} + O(h^4)$$

and

$$\cos\Gamma_2 - \varepsilon_2 = \cos\bar{\Gamma} + O(h^4).$$

Therefore,

$$\begin{aligned}
S &= \cos\Gamma + \cos\bar{\Gamma} + O(h^4) \\
&= 2\cos\Gamma + O(h^4)
\end{aligned}$$

and

$$\begin{aligned}
D &= 4\theta^2\frac{1}{2}h^2\alpha_1\Phi(\Gamma)^2\Gamma^{-1}\sin\Gamma\cdot\frac{1}{2}h^2\alpha_2\Phi(\bar{\Gamma})^2\bar{\Gamma}^{-1}\sin\bar{\Gamma} + O(h^6) \\
&= -h^4\beta^2\Phi(\Gamma)^2\Gamma^{-1}\Phi(\bar{\Gamma})^2\bar{\Gamma}^{-1}\sin^2\Gamma + O(h^6).
\end{aligned}$$

The result follows from Lemma 6. ∎

The proposition says that for all integers n and for all Γ that are not integer multiples of π we must have

$$\Phi(\Gamma)\Phi(2\pi n - \Gamma) = 0$$

to avoid type-1 instability. This can be achieved by constructing $\Phi(\Gamma)$ to vanish on $[\pi, 2\pi] \cup [3\pi, 4\pi] \cup \cdots$. If we satisfy this, then clearly $D \geq 0$ and type-1 instability is entirely avoided.

That LongAverage suffers from type-1 instability can be seen by choosing $\Gamma = \frac{2}{3}\pi$ and $\bar{\Gamma} = \frac{4}{3}\pi$, for which the stability matrix has spectral radius

$$1 + h^2|\beta|\frac{81}{128\sqrt{2}\pi^3}.$$

The next proposition is the result of an attempt to find the worst possible case of type-2 instability. First a lemma is needed.

LEMMA 7. *If $D \geq 0$ but $|S| > 2 - \sqrt{D}$, then the largest root of the LHS of* (28) *has modulus*

$$1 + \frac{1}{2}\delta + \sqrt{\delta + \frac{1}{4}\delta^2}$$

where

$$\delta = |S| - (2 - \sqrt{D}).$$

Proof. The LHS of (28) factors as

$$(\lambda^2 - (S + \sqrt{D})\lambda + 1)(\lambda^2 - (S - \sqrt{D})\lambda + 1).$$

The root λ of maximum modulus is a root of

$$\lambda^2 - (S + \mathrm{sign}(S)\sqrt{D})\lambda + 1 = \lambda^2 - \mathrm{sign}(S)(2 + \delta)\lambda + 1,$$

from which the lemma follows. ∎

PROPOSITION 2. *Let*

$$\Gamma_1 = \Gamma - \frac{1}{2}h^2\alpha_1\Phi(\Gamma)^2\Gamma^{-1},$$

where Γ is an odd integer multiple of π, and let Γ_2 not be an odd integer multiple of π. Then the spectral radius of the error propagation matrix is

$$1 + \frac{1}{2}h^2\alpha_1\Phi(\Gamma)^2\Gamma^{-1} + O(h^4).$$

Proof. Letting $\Delta = \frac{1}{2}h^2\alpha_1\Phi(\Gamma)^2\Gamma^{-1}$, we have

$$\begin{aligned}
\Gamma_1 &= \Gamma - \Delta, \\
\varepsilon_1 &= (\Delta + O(h^4))\sin(\Gamma - \Delta) \\
&= \Delta^2 + O(h^6), \\
\cos\Gamma_1 &= -1 + \frac{1}{2}\Delta^2 + O(h^8),
\end{aligned}$$

so

$$\cos\Gamma_1 - \varepsilon_1 = -1 - \frac{1}{2}\Delta^2 + O(h^6).$$

Therefore

$$\begin{aligned}
S &= \cos\Gamma_2 - 1 - \varepsilon_2 - \frac{1}{2}\Delta^2 + O(h^6), \\
|S| &= 1 - \cos\Gamma_2 + \varepsilon_2 + \frac{1}{2}\Delta^2 + O(h^6), \\
D &= \left(\cos\Gamma_2 - \varepsilon_2 - \left(-1 - \frac{1}{2}\Delta^2 + O(h^6)\right)\right)^2 + O(h^6),
\end{aligned}$$

and

$$\sqrt{D} = \cos\Gamma_2 - \varepsilon_2 + 1 + \frac{1}{2}\Delta^2 + O(h^6).$$

The result follows from Lemma 7. ∎

To avoid type-2 instability, it is necessary to have $\Phi(\Gamma)$ vanish at odd integer multiples of π.

The next proposition is the result of an attempt to find the worst possible case of type-3 instability.

PROPOSITION 3. *Let*

$$\Gamma_1 = \Gamma - \frac{1}{2}h^2\alpha_1\Phi(\Gamma)^2\Gamma^{-1},$$

where Γ is a positive even integer multiple of π, and let Γ_2 not be an even integer multiple of π. Then the spectral radius of the error propagation matrix is

$$1 + \frac{1}{2}h^2\alpha_1\Phi(\Gamma)^2\Gamma^{-1} + O(h^4).$$

Proof. The proof is very similar to that of Proposition 2. ∎

To avoid type-3 instability, it is necessary to have $\Phi(\Gamma)$ vanish at positive even integer multiples of π.

Next we show that LongAverage does not possess instabilities of types 2 and 3.

PROPOSITION 4. *LongAverage possesses neither type-2 nor type-3 instability provided that $h\omega \le 2$ where ω^2 is the spectral radius of U_{qq}.*

First we prove a lemma.

LEMMA 8.

$$\left|\frac{\sin\Gamma}{\Gamma}\right|^3 \le \frac{1}{2}(\cos\Gamma + 1) \quad \textit{and} \quad \left(\frac{\sin\Gamma}{\Gamma}\right)^3 \ge \frac{1}{4}(\cos\Gamma - 1).$$

Proof. Let $\psi = \frac{1}{2}\Gamma$ and this becomes

1. $\left|\dfrac{\sin 2\psi}{2\psi}\left(\dfrac{\sin\psi}{\psi}\right)^2\right| \le 1$, and
2. $-\dfrac{1}{2}\psi^2 \le \dfrac{\sin 2\psi}{2\psi}\cos^2\psi$.

The first is clearly true. The second is true for $\psi \le \pi/2$ because the RHS ≥ 0 and true for $\psi \ge \pi/2$ because the LHS ≤ -1. ∎

Proof of Proposition 4. Assume $h\omega \le 2$. The spectral radius of U_{qq} is

$$\omega^2 = \frac{\alpha_1 + \alpha_2}{2} + \sqrt{\left(\frac{\alpha_1 - \alpha_2}{2}\right)^2 + \beta^2},$$

so $\alpha_i \le \omega^2 \le 4h^{-2}$ and

$$\beta^2 \le (4h^{-2} - \alpha_1)(4h^{-2} - \alpha_2).$$

The upper bound on α_i implies that ε_i lies in the closed interval from 0 to $2(\sin\Gamma_i/\Gamma_i)^3$. Combining this with Lemma 7, we conclude

$$\frac{1}{2}(\cos\Gamma_i - 1) \le \varepsilon_i \le \cos\Gamma_i + 1. \tag{32}$$

Assume $D \ge 0$. We need to show $\sqrt{D} \le 2 - |S|$, for which it is enough to show

1. $|S| \le 2$,
2. $D \le (2 - S)^2$,
3. $D \le (2 + S)^2$.

Substitute into the definitions of S and D given in Lemma 4 and we are left with having to show

1. $-2 \le \cos\Gamma_1 - \varepsilon_1 + \cos\Gamma_2 - \varepsilon_2 \le 2$,
2. $\theta^2\varepsilon_1\varepsilon_2 \le (1 - \cos\Gamma_1 + \varepsilon_1)(1 - \cos\Gamma_2 + \varepsilon_2)$,
3. $\theta^2\varepsilon_1\varepsilon_2 \le (1 + \cos\Gamma_1 - \varepsilon_1)(1 + \cos\Gamma_2 - \varepsilon_2)$.

To show item 1, it is enough to show

$$-1 \le \cos\Gamma_i - \varepsilon_i \le 1,$$

which follows from (32). To show item 2, it is enough to show

$$|\varepsilon_i| \le 1 - \cos\Gamma_i + \varepsilon_i,$$

for which we need show only that

$$-\varepsilon_i \le 1 - \cos\Gamma_i + \varepsilon_i,$$

which again follows from (32). To show item 3, we note that

$$\begin{aligned}\theta^2\varepsilon_1\varepsilon_2 &= \frac{h^4\beta^2}{4}\left(\frac{\sin\Gamma_1 \sin\Gamma_2}{\Gamma_1\Gamma_2}\right)^3 \\ &\le \left|\left(2 - \frac{h^2\alpha_1}{2}\right)\left(\frac{\sin\Gamma_1}{\Gamma_1}\right)^3\left(2 - \frac{h^2\alpha_2}{2}\right)\left(\frac{\sin\Gamma_2}{\Gamma_2}\right)^3\right|,\end{aligned}$$

and hence it suffices to show

$$\left|2\left(\frac{\sin\Gamma_i}{\Gamma_i}\right)^3 - \varepsilon_i\right| \le 1 + \cos\Gamma_i - \varepsilon_i.$$

This is equivalent to showing that

$$2\left(\frac{\sin\Gamma_i}{\Gamma_i}\right)^3 \le 1 + \cos\Gamma_i$$

and

$$(h^2\alpha_i - 2)\left(\frac{\sin\Gamma_i}{\Gamma_i}\right)^3 \le 1 + \cos\Gamma_i.$$

Both of these follow from Lemma 8, using the fact that $|h^2\alpha_i - 2| \le 2$. ∎

It follows from the analyses given that if we are using spectral methods, we can easily create a perfect filter—one which vanishes or is very small for $h\Omega$ on $[\pi, 2\pi] \cup [3\pi, 4\pi] \cup \cdots$. The corresponding basis function is a multiple of its Fourier cosine transform and unfortunately probably does not have compact support.

REFERENCES

[1] J. J. Biesiadecki and R. D. Skeel, *Dangers of multiple-time-step methods*, J. Comput. Phys., 109 (1993), pp. 318–328.

[2] T. Bishop, R. D. Skeel, and K. Schulten, *Difficulties with Multiple Timestepping and the Fast Multipole Algorithm in Molecular Dynamics*, manuscript.

[3] M. Fixman, *Simulation of polymer dynamics.* I. *General theory*, J. Chem. Phys., 69 (1978), pp. 1527–1537.

[4] T. Forester and W. Smith, *On multiple time-step algorithms and the Ewald sum*, Mol. Sim., 13 (1994), pp. 195–204.

[5] H. Grubmüller, *Dynamiksimulation sehr großer Makromoleküle auf einem Parallelrechner*, master's thesis, Physik-Dept. der Tech. Univ. München, Munich, 1989.
[6] H. Grubmüller, *Molekular dynamik von Proteinen auf langen Zeitskalen*, Ph.D. thesis, Physik-Dept. der Tech. Univ. München, Munich, 1994.
[7] H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten, *Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions*, Mol. Sim., 6 (1991), pp. 121–142.
[8] H. Grubmüller and P. Tavan, *Efficient algorithms for molecular dynamics simulations of proteins: How good are they?*, 1994, manuscript.
[9] D. D. Humphreys, R. A. Friesner, and B. J. Berne, *A multiple-time-step molecular dynamics algorithm for macromolecules*, J. Phys. Chem., 98 (1994), pp. 6885–6892.
[10] M. A. López-Marcos, J. M. Sanz-Serna, and R. D. Skeel, *Explicit symplectic integrators using Hessian–vector products*, SIAM J. Sci. Comput., 18 (1997), pp. 223–238.
[11] G. Rowlands, *A numerical algorithm for Hamiltonian systems*, J. Comput. Phys., 97 (1991), pp. 235–239.
[12] J. Sanz-Serna and M. Calvo, *Numerical Hamiltonian Problems*, Chapman and Hall, London, 1994.
[13] J. Sanz-Serna and A. Portillo, *Classical numerical integrators for wave-packet dynamics*, J. Chem. Phys., 104 (1996), pp. 2349–2355.
[14] R. D. Skeel and J. J. Biesiadecki, *Symplectic integration with variable stepsize*, Ann. Numer. Math., 1 (1994), pp. 191–198.
[15] M. Tuckerman, B. J. Berne, and G. J. Martyna, *Reversible multiple time scale molecular dynamics*, J. Chem. Phys., 97 (1992), pp. 1990–2001.
[16] M. Watanabe and M. Karplus, *Simulation of macromolecules by multiple-time-step methods*, J. Phys. Chem., 99 (1995), pp. 5680–5697.
[17] J. Wisdom, *The origin of the Kirkwood gaps: A mapping for asteroidal motion near the* 3/1 *commensurability*, Astrophysics J., 87 (1982), pp. 577–593.
[18] J. Wisdom, M. Holman, and J. Touma, *Symplectic correctors*, in Integration Algorithms and Classical Mechanics, J. E. Marsden, G. W. Patrick, and W. F. Shadwick, eds., Fields Instit. Commun. 10, American Mathematical Society, Providence, RI, 1996, pp. 217–244.
[19] D. Xu, K. Schulten, O. M. Becker, and M. Karplus, *Temperature quench echoes in proteins*, J. Chem. Phys., 103 (1995), pp. 3112–3123.