Integration Schemes for Molecular Dynamics and Related Applications[1]

Robert D. Skeel
Department of Computer Science (and Beckman Institute),
University of Illinois,
Urbana,
Illinois 61801,
USA

Abstract

Presented are a variety of modern practical techniques for the derivation of integration schemes that are useful for molecular dynamics and a variety of related applications. In particular, the emphasis is on Hamiltonian systems, including those with constraints, and to a lesser extent stochastic differential equations. Among the techniques discussed are operator splitting, multiple time stepping, and accuracy enhancement through "post-processing." Attention is also given to analytical tools for selecting among different integration schemes, for example, small-time-step analysis of the backward error, linear analysis, and small-energy analysis.

1 Introduction

Presented are a variety of modern practical techniques for the derivation of numerical integration schemes that are useful for classical mechanical molecular dynamics (MD) and related applications. This is not a comprehensive survey but more of a sampling of issues in the construction of numerical integrators. The emphasis is on Hamiltonian systems, including those with constraints, and to a less extent stochastic differential equations. This material is presented here in the belief that techniques developed for a particular application are invariably useful in other domains and can often be abstracted and applied to generic problems.

Three aspects of molecular dynamics deserve special mention: (i) the chaotic nature of the differential equations, (ii) the large number of degrees of freedom, and (iii) the long and multiple time scales. These are discussed in the next three paragraphs.

The topic of this article might be considered to be numerical integrators for chaotic ordinary differential equations (ODEs). Chaos is a feature not only of molecular dynamics but also of galaxy simulations and climate modeling. For longer integrations, accurate trajectories are not expected. Wanted is statistically correct behavior and sampling of phase space. More specifically, initial velocities are randomly generated, and it is the evolution of the probabilities and their steady-state values that must be accurate. These probabilities are derived from the probability density function of the phase space variables, which satisfy an underlying linear homogeneous hyperbolic system of partial differential equations (in very many independent variables). Parts of this formulation are slightly speculative and nonrigorous. They depend on the hypothesis that the dynamical system is "mixing" (although it might suffice merely to be ergodic).

It is routine to simulate biomolecular systems with tens of thousands of atoms but desirable to simulate millions of atoms. Parallelism is an effective though expensive way to cope with large system size. Less expensive is to use Langevin dynamics, in which most degrees of freedom are omitted, and the missing forces are replaced by averaged and stochastic forces. Typically, the neglected degrees of freedom are those due to solvent molecules.

It is routine to simulate MD for hundreds of thousands of time steps but processes of biological interest are often on time scales that are yet millions of time longer. The highest frequency components of the dynamics are of small amplitude and of generally no intrinsic interest. Hence, it is reasonable to regard the equations to be "stiff-oscillatory" and longer time steps to be the major computational challenge. Only modest progress has been made in extending the length of the time step. A factor of two can be obtained by imposing constraints that remove the restraining effects of bond-length vibrations. A factor of four is possible with the use of multiple time steps to evaluate the more distant forces less often. For the Langevin equations, which contains friction forces, greater gains can been achieved. The friction inhibits instability and permits multiple time stepping to achieve its full potential. Also, the infinite friction limit sometimes provides an adequate approximation, reducing the second order Langevin equations to the first order stochastic ODEs of Brownian dynamics.

A secondary theme of this article concerns the subtleties of the numerical integration of special second order ODEs $(\mathrm{d}^2/\mathrm{d}t^2)q = F(q)$. Both the stability and the accuracy issues are special.

It is common for special second order ODEs to have "neutral" linear stability thereby requiring a more delicate, nonlinear stability analysis, necessitating excursions into dynamical systems theory. Hamiltonian systems of ODEs are a special case, for which it is desirable that the integrator be symplectic. The most successful technique for constructing effective symplectic integrators and preserving other qualities of the original problem is the use of operator splitting. Also, for problems as sensitive as these, the effects of finite precision must be considered. Happily, there are very simple ways to maintain the quality of the numerical solution.

Because the first derivative $(\mathrm{d}/\mathrm{d}t)q$ is not needed for calculating the "force" $F(q)$, it is a somewhat arbitrary decision as to how to define the numerical approximation to the first derivative. This leads to many different forms of the same method and to uncertainty regarding the accuracy of the method. The most illuminating way of understanding accuracy, especially for Hamiltonian systems, is by means of the "backward" error, usually as an asymptotic expansion in powers of the time step Δt. Also of value, particularly for special second order ODEs, is the concept of "effective" accuracy, in which the analyst does not impose an *a priori* interpretation on the meaning of values generated by a numerical integrator. Indeed, it is very practical to obtain the full effective accuracy of a numerical trajectory by "post-processing."

An outline follows:

2 Newtonian Dynamics

The Hamiltonian systems that commonly model classical molecular dynamics are chaotic. Underlying these unstable systems of ODEs are *stable* partial differential equations that describe the evolution of the probability density function for ensembles of trajectories produced by the system of ODEs. Calculating ensemble averages is the goal of MD simulations.

In a classical Newtonian mechanics model, each atom is a particle with position $\mathbf{r}_i$ and mass m_i. These positions are determined by the second order ODEs

$$m_i \frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathbf{r}_i = -\nabla_i U(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \tag{1}$$

for $i = 1, 2, \ldots, N$ where the potential energy function

$$U(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \text{bonded interactions} + \sum_{j>i}^{N} U_{ij}^{\text{nonbonded}}.$$

The first of the two terms consists of 2-, 3-, and 4-body forces. The second term consists of electrostatic interactions and the van der Waals attraction and hard-core repulsion. The electron distribution is accounted for by assigning partial charges to atoms. A third term would be present if "hydrogen bonds" are treated explicitly. These force fields are derived from (i) *ab initio* quantum mechanics calculations, (ii) spectroscopic measurements of small molecules that are similar to parts

of macromolecules, and (iii) fitting to measured constitutive properties such as diffusion coefficients and dielectric constants. Well known force fields include CHARMM, AMBER, GROMOS, and OPLS. Books on these and other aspects of molecular dynamics include [1, 33, 43, 65]. An article on MD integrators is [68].

An important matter is the treatment of boundaries. A simple treatment is to surround the system with a thick layer of solvent molecules in a vacuum. To prevent the occasional solvent molecule from evaporating, artificial restraining forces are often imposed. Alternatively, an implicit solvent model is used yielding a system of Langevin equations (discussed in section 8). However, the most popular method is to impose periodicity on the system.

2.1 Properties

Let $q = (\mathbf{r}_1^{\mathrm{T}}, \mathbf{r}_2^{\mathrm{T}}, \ldots, \mathbf{r}_N^{\mathrm{T}})^{\mathrm{T}}$ and defining a collective momenta vector $p = (m_1\mathbf{v}_1^{\mathrm{T}}, m_2\mathbf{v}_2^{\mathrm{T}}, \ldots, m_N\mathbf{v}_N^{\mathrm{T}})^{\mathrm{T}}$ where $\mathbf{v}_i = (\mathrm{d}/\mathrm{d}t)\mathbf{r}_i$ are velocities. Then the system (1) can be written

$$\frac{\mathrm{d}}{\mathrm{d}t}q = M^{-1}p, \quad \frac{\mathrm{d}}{\mathrm{d}t}p = F(q)$$

where M is a diagonal matrix of masses (each mass replicated three times) and the collective force vector $F(q) = -U_q(q)$. These equations are a special case of a Hamiltonian system of differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t}q = H_p(q,p), \quad \frac{\mathrm{d}}{\mathrm{d}t}p = -H_q(q,p). \tag{2}$$

with Hamiltonian $H(q,p) = K(p) + U(q)$ where the kinetic energy $K(p) = \frac{1}{2}p^{\mathrm{T}}M^{-1}p$.

Propagating the solution $(q(t), p(t))$ of a Hamiltonian system (2) for a fixed time increment Δt defines a transformation, the flow. Let $\phi_{\Delta t H}$ denote the Δt-flow for the system with Hamiltonian H:

$$\phi_{\Delta t H}(q(t), p(t)) = (q(t+\Delta t), p(t+\Delta t)).$$

It is valid to use the product $\Delta t H$ as the subscripted argument (without a comma) because the Δt-flow for Hamiltonian H is simply the 1-flow for Hamiltonian $\Delta t H$.

Hamiltonian systems have special properties, one of which is that the flow is a *symplectic* map. A transformation $\bar{y} = X(y)$ from phase space to itself where $y = (q,p)$ is *symplectic* if its Jacobian matrix X_y is symplectic:

$$X_y^{\mathrm{T}} J X_y = J$$

where

$$J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

(Note that a symplectic matrix is very easy to invert.) An important consequence of the symplectic property is that volume in phase space remains invariant under such a transformation.

Because $H(q,-p) = H(q,p)$, the Hamiltonian system possesses the property of time reversibility, which means that we can trace back the trajectory simply by reversing the momenta and integrating the same differential equation.

Linear and angular momentum are conserved if the potential is invariant under translation and rotation:

$$U(\{Q\mathbf{r}_j + \mathbf{d}\}) = U(\{\mathbf{r}_j\}).$$

The energy $H(q,p)$ is always conserved.

Letting $\phi = \phi_{\Delta t H}$, the properties given in the last three paragraphs can all be expressed in the form $\mathcal{F}(\phi)(y) = \mathcal{F}(\mathrm{id})(y)$ for some operator $\mathcal{F}$:

1. symplecticness:

$$\phi_y(y)^{\mathrm{T}} J \phi_y(y) = J.$$

2. reversibility:

$$R\phi(R\phi(y)) = y$$

where $R = \mathrm{diag}(I, -I)$.

3. conservation laws:

$$\begin{aligned}
&\text{linear momentum} \quad a_i^{\mathrm{T}} \phi(y) = a_i^{\mathrm{T}} y, \quad i = 1, 2, 3, \\
&\text{angular momentum} \quad \phi(y)^{\mathrm{T}} A_i \phi(y) = y^{\mathrm{T}} A_i y, \quad i = 1, 2, 3, \\
&\text{energy} \quad H(\phi(y)) = H(y).
\end{aligned}$$

Which of these are preserved by numerical integrators?

2.2 The Liouville equation

The chaotic nature of the atomic trajectories is pronounced: a roundoff error introduced into the computations will after a short time overwhelm the theoretically correct trajectory, even with the use of double precision. Over a 50-picosecond duration growth is by a factor 10^{15} [1].

Appropriate accuracy criteria depend on the purpose of the computation:

1. exploration of phase space [65, p. 434]. MD is one of many ways to search for conformations of a biomolecules having minimal or near minimal free energy.

2. sampling of phase space. The goal is to sample ensembles of molecular systems from some given probability distribution and to compute averages of certain quantities or to find structures having high probabilities. Choosing suitably random initial conditions is difficult for position variables, so the sampling is preceded by an equilibration phase that randomizes the initial conditions. For the sampling phase, Monte Carlo methods are popular for small molecules but they suffer from high rejection rates in the case of macromolecules, where MD and hybrid methods are more often used. The use of one long MD simulation for sampling relies on the ergodic hypothesis that the time average is equal to the ensemble average.

3. accurate short-time trajectories. The calculation of time correlation functions requires a collection of trajectories with initial conditions drawn from an appropriate ensemble. One long MD trajectory can be used for this purpose.

4. statistically accurate long-time behavior [65, p. 354]. The intent is to simulate the dynamics of ensembles of molecular systems, with initial conditions chosen at random from an appropriate distribution. The simulation should give the correct probabilities of making a transition from one structure to another after some designated time interval. It is typical practice for biomolecules to simulate only a small number of instances. However, this may not be a serious matter because, unlike general macromolecules (such as polymers from which materials are constructed), biomolecules are designed to function in a fairly consistent and reproducible manner on longer time scales even though they experience thermal fluctuations on short time scales.

Let us formalize a probabilistic accuracy criterion for constant energy dynamics. Assume that the initial conditions are drawn from a distribution with probability density (function) $\rho_0(y)$. Imagine generating an "ensemble" of trajectories from these initial conditions. Let $\rho(y,t)$ be the probability density that results. The quantities of interest are statistical and generally involve integrals of the form

$$\int g(y)\rho(y,t)\mathrm{d}y.$$

(Structure determination can also be formulated in this way.) It is quantities such as these that are required to be accurate. Because of the approximate nature of the force fields, it is necessary that these quantities be insensitive in some sense to small changes in the energy function. (Otherwise, molecular dynamics as a methodology is on shaky ground.)

For the ODE system $(\mathrm{d}/\mathrm{d}t)y = f(y)$, the probability density $\rho(y,t)$ for an ensemble of trajectories obeys the Liouville equation

$$\rho_t + \nabla^{\mathrm{T}}(\rho f(y)) = 0$$

where ∇^{T} denotes the divergence operator. This is a homogeneous linear hyperbolic partial differential equation. The solutions of $(\mathrm{d}/\mathrm{d}t)y = f(y)$ constitute the characteristics of this partial differential equation. If $\nabla^{\mathrm{T}} f(y) = 0$, as it is for Hamiltonian systems, where $f(y) = JH_y(y)$, this becomes

$$\rho_t + f(y)^{\mathrm{T}}\rho_y = 0.$$

For the special Hamiltonian $H(q,p) = \frac{1}{2}p^{\mathrm{T}}M^{-1}p + U(q)$, we get

$$\rho_t + (M^{-1}p)^{\mathrm{T}}\rho_q - U_q(q)^{\mathrm{T}}\rho_p = 0.$$

Assume that initial conditions are restricted so that center of mass, linear momentum, and angular momentum are always zero and that the only nonzero conserved quantity is energy. Does $\rho(y,t)$ have a limit as $t \to \infty$? No, this would contradict reversibility. But it is conceivable that there is a limit in the weak sense:

$$\lim_{t\to\infty} \int g(y)\rho(y,t)\mathrm{d}y = \int g(y)\bar{\rho}(y)\mathrm{d}y$$

for any $g \in \mathrm{C}^\infty$. If there is such a limit, the dynamical system is said to be *mixing* [64, 43]. The meaning of this weak limit is that the difference $\rho(y,t) - \bar{\rho}(y)$ becomes more and more highly fluctuating as a function of y as $t \to \infty$. A yet weaker property is to be *ergodic*:

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \rho(y,\tau)\mathrm{d}\tau = \bar{\rho}(y).$$

In either case the equilibrium density must satisfy

$$(\bar{\rho}_y(y))^{\mathrm{T}} JH_y(y) = 0,$$

which implies that $\bar{\rho}(y)$ is a conserved quantity, which implies $\bar{\rho}(y) = \text{function}(H(y))$. Hence $\bar{\rho}(y)$ is essentially uniform in a very thin energy shell $E \le H(y) \le E + \epsilon$. Letting $\epsilon \to 0$, we get

$$\bar{\rho}(y) \propto \frac{1}{\|H_y(y)\|_2} \qquad \text{for } H(y) = E.$$

If the density almost always tends to some limit density $\bar{\rho}(y)$ that is bounded away from zero, this suggests that almost every trajectory passes arbitrarily close to every point on its energy surface (if we wait long enough). This ergodic property of trajectories justifies the use of a time average to calculate an ensemble average. For ensembles other than the microcanonical ensemble (constant energy), the Hamiltonian needs to be extended in some way. Examples are given in section 7.

For further discussion of mathematical issues and for references to the literature, the preprints [102, 29] are recommended.

3 The Leapfrog Method

The equations of motion are typically integrated by the leapfrog/Störmer[114]/Verlet method, which is reported [120] to date as far back as 1793 [21]. The method used by Verlet [125] was expressed in the form

$$M\frac{q^{n+1}-2q^n+q^{n-1}}{\Delta t^2}=F(q^n),\quad v^n=\frac{q^{n+1}-q^{n-1}}{2\Delta t} \tag{3}$$

where v^n is the collective velocity vector and $q^1=q^0+\Delta t v^0+(\Delta t^2/2)M^{-1}F(q^0)$. However, there are a surprising variety of other ways to write the leapfrog method. These and numerous other interesting but hardly well known facts about the leapfrog method will emerge as we use this scheme to illustrate a variety of techniques for analysis and for algorithm construction.

Analytical error analysis of numerical methods must necessarily make simplifications—if we knew precisely what the error was, we would also know the solution and have little need for the numerical method. Three types of simplifying assumptions are popular, each giving a different mode of error analysis:

1. error bounds. By making repeated worst case assumptions one can derive rigorous upper bounds on the magnitude of the error. These are generally much greater than the worst possible error. Nonetheless, they do convey qualitative information and are useful in comparing alternative methods. However, good error bounds are difficult to obtain. Overall, this approach is the least rewarding.

2. exact error analysis for simple test problems. Usually these are linear constant coefficient problems. Nonetheless, it is not difficult to relate the conclusions to real problems, and these conclusions generally have good predictive value.

3. asymptotic expansions in some small parameter. Typically this small parameter is Δt, but another possibility is to analyze small deviations from equilibrium. Conclusions from these analyses also have good predictive value.

In sections 3.2–3.4 we pursue the second and third of these for the leapfrog method.

Discretization error for leapfrog can be rather dramatically reduced by the use of post-processing discussed in section 3.5. Another kind of error, roundoff error, is treated in section 3.6.

3.1 Derivation

The second order difference equation formulation (3) given above is known to accelerate by a factor of n the accumulation of roundoff error. Here we derive a form of the method less affected by roundoff and more befitting of the name "leapfrog."

The exact calculation of the flow $\phi_{\Delta t(K+U)}$ is, of course, generally intractable. However, the special cases $\phi_{\Delta tK}$ and $\phi_{\Delta tU}$ have a very simple closed form. For example, $\phi_{(\Delta t/2)U}(q^n,p^n)$ can be obtained by solving an initial value problem for

$$\frac{\mathrm{d}}{\mathrm{d}t}q=0,\quad \frac{\mathrm{d}}{\mathrm{d}t}p=-U_q(q),$$

yielding

$$\phi_{(\Delta t/2)U}(q^n,p^n)=\begin{pmatrix} q^n \\ p^n-\frac{1}{2}\Delta t U_q(q^n)\end{pmatrix}.$$

This is the motivation for operator splitting. In particular, the leapfrog method is the result of a Strang splitting:

$$\begin{aligned}\phi^{\mathrm{LF}}_{\Delta t(K+U)} &= \phi_{(\Delta t/2)U} \circ \phi_{\Delta t K} \circ \phi_{(\Delta t/2)U} \\ &= \phi_{\Delta t(K+U)} + O(\Delta t^3).\end{aligned}$$

Over many steps these errors accumulate to yield trajectory errors that are, in fact, $O(\Delta t^2)$. That this composition of propagators should be a reasonable approximation, let alone of second order accuracy, may not be obvious. The justification is due to the exponential nature of solution propagators, a subject that is discussed further in appendix A. And the inaccuracy is due to the fact that the rule for a product of exponentials is not exact for operators that do not commute.

The leapfrog method can be expressed as follows: We begin the step with values q^n, p^n, and $F^n = F(q^n)$ obtained from the previous step. Then we compute

$\phi_{(\Delta t/2)U}$ (half-kick):

$$p^{n+1/2} = p^n + \frac{\Delta t}{2} F^n,$$

$\phi_{\Delta t K}$ (drift):

$$q^{n+1} = q^n + \Delta t M^{-1} p^{n+1/2},$$

$\phi_{(\Delta t/2)U}$ (half-kick):

$$\begin{aligned}F^{n+1} &= F(q^{n+1}), \\ p^{n+1} &= p^{n+1/2} + \frac{\Delta t}{2} F^{n+1}.\end{aligned}$$

(If we combine the last half kick of one step with the first half kick of the next step, we get a form of the method that uses p values only at midpoints of intervals. This form of the method is associated with the name "leapfrog." Alternatively, we can define q values at midpoints of intervals using $q^{n+1/2} = q^n + \frac{1}{2}\Delta t M^{-1} p^{n+1/2}$, $q^{n+1} = q^{n+1/2} + \frac{1}{2}\Delta t M^{-1} p^{n+1/2}$.) These equations exactly model the following situation: an impulse of $\frac{1}{2}\Delta t F^n$ at time t^n+ followed by an impulse of $\frac{1}{2}\Delta t F^{n+1}$ at time $t^{n+1}-$. Note that because the term $\frac{1}{2}\Delta t F^{n+1}$ is treated as an impulse just before time t^{n+1}, it has no effect on the value of q^{n+1} and the method is explicit. In fact, as noted by Wisdom [128], the leapfrog method, exactly satisfies the equations of motion

$$M \frac{\mathrm{d}^2}{\mathrm{d}t^2} q = \sum_n \Delta t \, \delta(t - n\Delta t) F(q).$$

There is an alternative form of the leapfrog method based on the splitting

$$\phi_{\Delta t(K+U)} = \phi_{(\Delta t/2)K} \circ \phi_{\Delta t U} \circ \phi_{(\Delta t/2)K} + O(\Delta t^3).$$

We can interpret this as half a drift followed by a kick followed by half a drift. We call this the *midpoint* form of the leapfrog method because the force is evaluated at the middle. The first form we call the *endpoint* form. We prefer the endpoint form because it has a more efficient generalization to multiple time stepping. The two forms are equivalent in the sense that there exists a coordinate transformation that relates numerical trajectories produced by each of them [112].

The flow $\phi^{\mathrm{LF}}_{\Delta t H}$ defined by the leapfrog method is symplectic. This is a consequence of two important facts. One is that Hamiltonian flows are symplectic. The other is that the composition of two symplectic transformations is also symplectic (as is the inverse of a symplectic transformation). Also, the leapfrog method preserves the conservation laws for linear and angular momentum and the property of time reversibility. One conservation law that is not maintained is conservation of energy. More is said about the significance of these properties in section 4.

3.2 Small-Δt analysis

At least three types of error expansions in powers of Δt are possible:

1. "global" error expansions. For example, time-reversible methods like the leapfrog method have an expansion

$$y^n = y(n\Delta t) + \Delta t^2 e_1(n\Delta t) + \Delta t^4 e_2(n\Delta t) + \cdots$$

containing only even powers of Δt. The identity of the coefficients $e_1(t)$, $e_2(t)$, ... in such an expansion is determined by substituting the expansion into the difference equation, expanding about $t = n\Delta t$, and equating like powers of Δt^2. One can by this type of analysis determine those more important terms of the error (the secular terms) that grow in time, e.g., [69]. This type of analysis has been fruitful in studying discretizations of periodic orbits. For such problems, symplectic integrators have been shown to yield global errors that increase only linearly as time t grows rather than quadratically as they do for non-symplectic integrators [15].

2. local error expansions. The local error for the leapfrog method is $\phi^{\mathrm{LF}}_{\Delta t H}(y(t)) - y(t + \Delta t)$. Because the first derivative v^n is not used, as q^n is, in an evaluation of a function, it hardly matters how v^n is defined. This arbitrariness is inherited by the definition of the local error whose value depends on how we choose to define v^n. However, the accuracy of the numerical approximation to q^n is not affected except by the manner in which the initial value $v(0)$ is incorporated into the solution. The arbitrariness can be avoided by treating the method as a second order difference approximation to a second order differential equation.

3. backward error expansions. The numerical solution can be expressed formally as the exact solution of an ODE with a perturbed right hand side expressed as an expansion in powers of Δt. This may be the most enlightening analysis. It was first developed in the context of computational fluid dynamics under the name "method of modified equations" [126]. It is interesting that a truncated asymptotic backward analysis yields a more accurate approximation to the numerical solution than does a truncated asymptotic expansion of the global error [48].

Here we explore only the third of these, namely, backward error analysis. To present the concept with the greatest clarity, we apply this technique first to the forward Euler approximation

$$\frac{y^{n+1} - y^n}{\Delta t} = f(y^n)$$

to the ODE system $(\mathrm{d}/\mathrm{d}t)y = f(y)$. The idea is to assume that $y^n = \tilde{y}(n\Delta t)$ where $\tilde{y}(t)$ satisfies a slightly different ODE system $(\mathrm{d}/\mathrm{d}t)\tilde{y} = \tilde{f}(\tilde{y})$ with $\tilde{f}$ assumed to possess an asymptotic expansion in powers of Δt. We begin with an expansion about some chosen point in time, $t = n\Delta t$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{y} + \frac{\Delta t}{2}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\tilde{y} + \frac{\Delta t^2}{6}\frac{\mathrm{d}^3}{\mathrm{d}t^3}\tilde{y} + \cdots = f(\tilde{y}).$$

Then we use the ODE $(\mathrm{d}/\mathrm{d}t)\tilde{y} = \tilde{f}(\tilde{y})$ and its derivatives $(\mathrm{d}^2/\mathrm{d}t^2)\tilde{y} = \tilde{f}_y(\tilde{y})\tilde{f}(\tilde{y})$, $(\mathrm{d}^3/\mathrm{d}t^3)\tilde{y} = \tilde{f}_{yy}(\tilde{y})\tilde{f}(\tilde{y})\tilde{f}(\tilde{y}) + \tilde{f}_y(\tilde{y})\tilde{f}_y(\tilde{y})\tilde{f}(\tilde{y})$, ... to get an equation involving only functions of $\tilde{y}(t)$ and not of its derivatives. (In general we have $(\mathrm{d}^k/\mathrm{d}t^k)\tilde{y}(t) = (\tilde{f}^{\mathrm{T}}\nabla)^k \mathrm{id}(\tilde{y}(t))$. See appendix A.) Moving all but the first term to the right-hand side and suppressing the dependence on $y = \tilde{y}(t)$, we get

$$\tilde{f} = f - \frac{\Delta t}{2}\tilde{f}_y\tilde{f} - \frac{\Delta t^2}{6}(\tilde{f}_{yy}\tilde{f}\tilde{f} + \tilde{f}_y\tilde{f}_y\tilde{f}) - \cdots.$$

We can develop an expansion for $\tilde{f}$ by successive substitution into the right-hand side. Initially take

$$\tilde{f} = f + O(\Delta t).$$

Then

$$\tilde{f} = f - \frac{\Delta t}{2} f_y f + O(\Delta t^2),$$

and then

$$\tilde{f} = f - \frac{\Delta t}{2} f_y f + \frac{\Delta t^2}{12}(f_{yy} f f + 4 f_y f_y f) + O(\Delta t^3),$$

etc.

We now do a backward analysis for the leapfrog method. By eliminating the intermediate stage in the leapfrog method, we get

$$\frac{p^{n+1} - p^n}{\Delta t} = \frac{1}{2}(F^n + F^{n+1}), \tag{4}$$

$$\frac{q^{n+1} - q^n}{\Delta t} = \frac{1}{2} M^{-1}(p^{n+1} + p^n) - \frac{\Delta t}{4} M^{-1}(F^{n+1} - F^n). \tag{5}$$

Expanding (5) and (4) in a Taylor series about the midpoint $t = (n + \frac{1}{2})\Delta t$, we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\tilde{q} + \frac{\Delta t^2}{24}\frac{\mathrm{d}^3}{\mathrm{d}t^3}\tilde{q} &= M^{-1}\tilde{p} + \frac{\Delta t^2}{8} M^{-1}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\tilde{p} + \frac{\Delta t^2}{4} M^{-1}\frac{\mathrm{d}}{\mathrm{d}t} U_q(\tilde{q}) + O(\Delta t^4), \\
\frac{\mathrm{d}}{\mathrm{d}t}\tilde{p} + \frac{\Delta t^2}{24}\frac{\mathrm{d}^3}{\mathrm{d}t^3}\tilde{p} &= -U_q(\tilde{q}) - \frac{\Delta t^2}{8}\frac{\mathrm{d}^2}{\mathrm{d}t^2} U_q(\tilde{q}) + O(\Delta t^4)
\end{aligned}$$

where $\tilde{q} = \tilde{q}((n + 1/2)\Delta t)$, $\tilde{p} = \tilde{p}((n + 1/2)\Delta t)$. Hence

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\tilde{q} &= M^{-1}\tilde{p} + \Delta t^2 \left(\frac{1}{8} M^{-1}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\tilde{p} + \frac{1}{4} M^{-1} U_{qq}(\tilde{q})\frac{\mathrm{d}}{\mathrm{d}t}\tilde{q} - \frac{1}{24}\frac{\mathrm{d}^3}{\mathrm{d}t^3}\tilde{q} \right) + O(\Delta t^4), \\
\frac{\mathrm{d}}{\mathrm{d}t}\tilde{p} &= -U_q(\tilde{q}) + \Delta t^2 \left(-\frac{1}{8} U_{qqq}(\tilde{q})(\frac{\mathrm{d}}{\mathrm{d}t}\tilde{q}, \frac{\mathrm{d}}{\mathrm{d}t}\tilde{q}) - \frac{1}{8} U_{qq}(\tilde{q})\frac{\mathrm{d}^2}{\mathrm{d}t^2}\tilde{q} - \frac{1}{24}\frac{\mathrm{d}^3}{\mathrm{d}t^3}\tilde{p} \right) + O(\Delta t^4),
\end{aligned}$$

where $U_{qq}(\tilde{q})$ is the Hessian of the potential energy function and a double summation involving $U_{qqq}(\tilde{q})$ is intended by the notation $(\frac{\mathrm{d}}{\mathrm{d}t}\tilde{q}, \frac{\mathrm{d}}{\mathrm{d}t}\tilde{q})$. By one application of successive substitution we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\tilde{q} &= M^{-1}\tilde{p} + \Delta t^2 \left(\frac{1}{6} M^{-1} U_{qq}(\tilde{q}) M^{-1}\tilde{p} \right) + O(\Delta t^4), \\
\frac{\mathrm{d}}{\mathrm{d}t}\tilde{p} &= -U_q(\tilde{q}) + \Delta t^2 \left(-\frac{1}{12} U_{qqq}(\tilde{q})(M^{-1}\tilde{p}, M^{-1}\tilde{p}) + \frac{1}{12} U_{qq}(\tilde{q}) M^{-1} U_q(\tilde{q}) \right) + O(\Delta t^4),
\end{aligned}$$

whence follows

$$\begin{aligned}
\tilde{H}(q,p) &= \frac{1}{2} p^{\mathrm{T}} M^{-1} p + U(q) \\
&\quad + \Delta t^2 \left(\frac{1}{12}(M^{-1}p)^{\mathrm{T}} U_{qq}(q) M^{-1} p - \frac{1}{24} U_q(q)^{\mathrm{T}} M^{-1} U_q(q) \right) + O(\Delta t^4).
\end{aligned} \tag{6}$$

The forgoing calculation can be done more efficiently using the formalism of Lie operators and the BCH formula, as described in appendix A.

3.3 Linear analysis

Normal mode analysis examines the dynamics in the vicinity of a potential energy minimum, for which the gradient

$$U_q(q^*) = 0$$

and the Hessian

$$U_{qq}(q^*) \geq 0,$$

meaning that it is semi-positive definite. The dynamics in a neighborhood of $q = q^*$ is described by

$$\frac{\mathrm{d}}{\mathrm{d}t}q = M^{-1}p, \quad \frac{\mathrm{d}}{\mathrm{d}t}p = -U_{qq}(q^*)(q - q^*)$$

if the $O(\|q - q^*\|^2)$ terms are neglected, which is the case for normal mode analysis. The resulting linear problem can be decoupled by a similarity transformation and with a translation and symplectic scaling of variables the system reduces to a set of harmonic oscillators

$$\frac{\mathrm{d}}{\mathrm{d}t}q_i = p_i, \quad \frac{\mathrm{d}}{\mathrm{d}t}p_i = -\omega_i^2 q_i \tag{7}$$

where the ω_i^2 are the eigenvalues of the mass-weighted Hessian $M^{-1/2}U_{qq}(q^*)M^{-1/2}$. Dropping the subscripts and writing in vector form as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} q \\ p \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega^2 & 0 \end{bmatrix}\begin{bmatrix} q \\ p \end{bmatrix},$$

we see that the solution is

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

If $\omega > 0$ is assumed, the variables can be scaled to yield a pure rotation by $-\omega t$ in phase space:

$$\begin{bmatrix} \omega q(t) \\ p(t) \end{bmatrix} = \begin{bmatrix} \cos\omega t & \sin\omega t \\ -\sin\omega t & \cos\omega t \end{bmatrix}\begin{bmatrix} \omega q(0) \\ p(0) \end{bmatrix}.$$

For leapfrog and other standard integrators applied to the original linear problem, precisely the same transformations decouple the difference equation to yield a leapfrog scheme applied to each of the harmonic oscillators (7). If $\omega\Delta t \geq 2$, the leapfrog method is unstable; if $\omega\Delta t < 2$, it is stable. In such a case, the Δt-flow is

$$\begin{bmatrix} \omega q^{n+1} \\ p^{n+1} \end{bmatrix} = D\begin{bmatrix} \cos\tilde{\omega}\Delta t & \sin\tilde{\omega}\Delta t \\ -\sin\tilde{\omega}\Delta t & \cos\tilde{\omega}\Delta t \end{bmatrix}D^{-1}\begin{bmatrix} \omega q^n \\ p^n \end{bmatrix} \tag{8}$$

where

$$\tilde{\omega} = \frac{2}{\Delta t}\arcsin\frac{\omega\Delta t}{2} = \omega(1 + O((\omega\Delta t)^2)$$

and

$$D = \mathrm{diag}\left(1, \sqrt{1 - \omega^2\Delta t^2/4}\right) = I + O((\omega\Delta t)^2).$$

Hence, we note that the leapfrog scheme imparts a "blue shift," to angular frequency ω, which is 1.5% for Δt equal to $\frac{1}{10}$ of a period.

Following is a derivation from [112] of the modified Hamiltonian for the harmonic oscillator. The leapfrog solution (8) has

$$\begin{bmatrix} \omega\tilde{q}(t) \\ \tilde{p}(t) \end{bmatrix} = D \begin{bmatrix} \cos\tilde{\omega}t & \sin\tilde{\omega}t \\ -\sin\tilde{\omega}t & \cos\tilde{\omega}t \end{bmatrix} D^{-1} \begin{bmatrix} \omega q^0 \\ p^0 \end{bmatrix}.$$

Differentiating with respect to t gives

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} \omega\tilde{q}(t) \\ \tilde{p}(t) \end{bmatrix} = \tilde{\omega} D J \begin{bmatrix} \cos\tilde{\omega}t & \sin\tilde{\omega}t \\ -\sin\tilde{\omega}t & \cos\tilde{\omega}t \end{bmatrix} D^{-1} \begin{bmatrix} \omega q^0 \\ p^0 \end{bmatrix},$$

and eliminating initial values from this and the previous equation gives the ODE system

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} \tilde{q}(t) \\ \tilde{p}(t) \end{bmatrix} = \tilde{\omega} J \mathrm{diag}\left(\omega(1-\omega^2\Delta t^2/4)^{1/2}, \omega^{-1}(1-\omega^2\Delta t^2/4)^{-1/2}\right) \begin{bmatrix} \tilde{q}(t) \\ \tilde{p}(t) \end{bmatrix}.$$

This is a Hamiltonian system arising from

$$\tilde{H}(q,p) = \frac{1}{\omega\Delta t}\arcsin\frac{\omega\Delta t}{2}((1-\omega^2\Delta t^2/4)^{-1/2}p^2 + (1-\omega^2\Delta t^2/4)^{1/2}\omega^2 q^2),$$

valid for $\omega\Delta t < 2$. The "unmodified" Hamiltonian is $\frac{1}{2}(p^2+\omega^2 q^2)$.

3.4 Small-energy analysis

Here we consider how nonlinearity affects behavior in the neighborhood of a stable equilibrium, for which

$$p^* = 0, \quad U_q(q^*) = 0, \quad \text{and} \quad U_{qq}(q^*) > 0,$$

meaning that it is positive definite. Without neglecting any higher powers of $\|q-q^*\|$, it is possible to make exact mathematical statements about the stability of a symplectic map in a "sufficiently small" neighborhood of a fixed point y^*. Unstable behavior in this situation would dictate the use of a smaller time step or a different integrator. Stability for an integrator means bounded growth in the error uniformly for all time for initial values in some neighborhood of its fixed point y^*. More precisely, let $y^{n+1} = \tilde{\phi}_{\Delta t H}(y^n)$ be a numerical integrator with fixed point $y^* = \tilde{\phi}_{\Delta t H}(y^*)$. The integrator is stable at y^* if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\|y^0 - y^*\| < \delta \qquad \Rightarrow \qquad \|y^n - y^*\| < \varepsilon \quad \text{for all } n.$$

Expanding the numerical flow $\tilde{\phi}_{\Delta t H}(y^n)$ about $y^n = y^*$, we can write

$$y^{n+1} - y^* = Q\,(y^n - y^*) + O(\|y^n - y^*\|^2)$$

where Q is the Jacobian matrix of the map at y^*.

Consider the scalar case. Let us assume that Δt is less than the linear stability threshold $2/\omega$. Then, as can be seen from eq. (8), Q has eigenvalues $\exp(\pm \mathrm{i}\tilde{\omega}\Delta t)$. Stability at equilibrium of the nonlinear map can be determined using the Moser twist theorem [4, p. 411]. If $\exp(\mathrm{i}\tilde{\omega}\Delta t)$ is not a root of unity or equivalently if the ratio of $\tilde{\omega}\Delta t$ to 2π is irrational, the map is stable. If $\exp(\mathrm{i}\tilde{\omega}\Delta t)$ is an nth root of unity, we have nth order *resonance* and the possibility of instability [4, pp. 390–398]. Third order resonances are normally unstable in the sense that stability is possible only if a certain coefficient in some expansion happens to vanish. Fifth and higher order resonances are normally stable in the sense that instability is possible only if a certain coefficient happens to vanish. Both

stability and instability are normal for fourth order resonances. The specifics depend on both the integrator and the function $U(q)$. For this reason both third and fourth order resonances should be regarded as generally unacceptable. In the case of the leapfrog method, $\tilde{\omega}\Delta t = \frac{2}{3}\pi$ (3:1 resonance) corresponds to $\omega\Delta t = \sqrt{3}$ and $\tilde{\omega}\Delta t = \frac{1}{2}\pi$ (4:1 resonance) corresponds to $\omega\Delta t = \sqrt{2}$. For rotations $\tilde{\omega}\Delta t$ near an unstable resonance, the stable neighborhood of y^* is very small. These positive results on stability were first elucidated in [99]. The article [79] explains that stable resonances induce oscillations with an amplitude that depends on the order of the resonance. Empirical evidence of instability due to third order resonance was first observed for the implicit midpoint method [72], which is unconditionally stable for linear problems.

The situation in higher dimensions is less conclusive. Assume the eigenvalues λ_1, λ_2, ..., λ_d of the Jacobian matrix Q of the map are all of unit modulus. If there are integers k_1, k_2, ..., k_d such that $\lambda_1^{k_1}\lambda_2^{k_2}\cdots\lambda_d^{k_d} = 1$, we have a resonance of order $n = |k_1|+|k_2|+\cdots+|k_d|$. In this more general case it is possible to have instability even if unstable resonances are avoided. But this instability will be of a less virulent type known as *Arnol'd diffusion*. Instability in higher dimensions has been observed for the implicit midpoint rule applied to models in structural mechanics [106, 39]. There are a large number of different frequencies present in a vibrating structure, so third order resonance is a plausible explanation for instability. In MD, resonance-induced instability has been demonstrated for the implicit midpoint rule applied to a biomolecule [101]. The same article proposes other implicit methods that avoid third and even fourth order resonances. Resonances can, in fact, be observed during normal MD simulations using the leapfrog method. Shown in Fig. 1 is plot of energy as a function of time that was produced during the course of testing the correctness

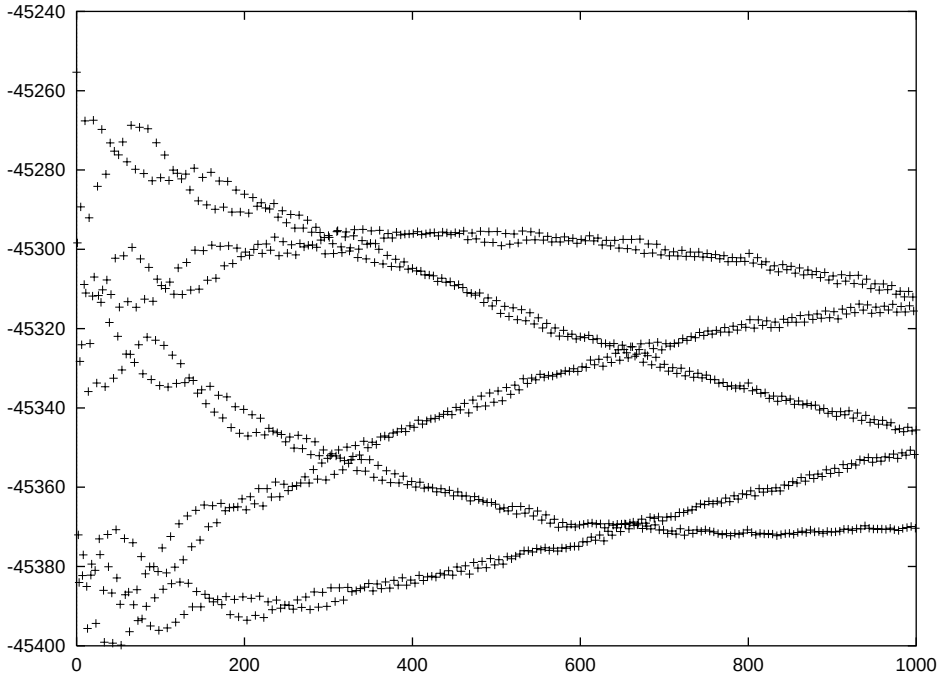

Figure 1: Energy as a function of time for a leapfrog simulation of MD.

of a simulation program, NAMD 1.5 [83], on a protein–water system. Clearly evident in the graph is 10th order resonance, which is not surprising since a time step of 1 fs is about one tenth the period of the fastest normal mode in the simulation. Because typical dynamical simulations involve a spectrum of different frequencies, it seems advisable to select the time step Δt for the leapfrog

method so that the rotation $\tilde{\omega}\Delta t$ per step is less than $\pi/2$. Hence, *the stability condition for the leapfrog method is effectively*

$$\omega\Delta t < \sqrt{2}$$

rather than $\omega\Delta t < 2$. This, in fact, is consistent with an observed stability limit of 2 fs or less for MD simulations

3.5 Effective accuracy and post-processing

It has been observed first by [14] and later in the context of Hamiltonian systems by [95, 118, 129, 78, 69] that the "intrinsic" accuracy of an integrator may be greater than what it seems—a slightly different interpretation of the numerical solution (q^n, p^n) may make the method look more accurate, e.g., as assessed by local error or backward error analysis. More specifically, the standard interpretation is that $y^n \approx y(n\Delta t)$ and a more favorable interpretation may be that

$$y^n \approx \hat{y}(n\Delta t) = X(y(n\Delta t))$$

where X is a coordinate transformation in phase space only slightly different from the identity mapping id. In other words we associate with a method a transformation X^{-1} that obtains more desirable values $X^{-1}(y^n)$ from the raw values y^n used by the integration scheme. Clearly, there are limits to what can be achieved, and, in particular, there is no transformation that removes the growing secular part of the error. In the case of the leapfrog method applied to the harmonic oscillator discussed previously, the error $D - I$ that causes the energy to oscillate is completely removable; whereas the error $\tilde{\omega} - \omega$ that causes a growing phase error is not at all removable.

A simple yet dramatic example of this phenomenon is the method [104], [37, p. 24]

$$p^{n+1} = p^n - \Delta t U_q(q^n), \quad q^{n+1} = q^n + \Delta t M^{-1} p^{n+1} \tag{9}$$

and its twin

$$q^{n+1} = q^n + \Delta t M^{-1} p^n, \quad p^{n+1} = p^n - \Delta t U_q(q^{n+1}), \tag{10}$$

either of which might be called the *symplectic Euler method*. As noted by Suzuki [118], either of (9) or (10) is equivalent to the leapfrog method. In particular, perform the change of coordinates

$$\begin{aligned} q &= \bar{q}, \\ p &= \bar{p} + \frac{\Delta t}{2} U_q(\bar{q}), \end{aligned}$$

and the first-order accurate method (9) becomes the second-order accurate leapfrog method for the new coordinates $\bar{q}^n, \bar{p}^n$. Note that $X = \mathrm{id} + O(\Delta t)$ and that X is symplectic.

In general, a transformation X such that a method better propagates $X(y(t))$ than it does $y(t)$ can be used to create a better numerical trajectory as follows:

$$\begin{aligned} y^n &= X(\bar{y}^n), \\ y^{n+1} &= \tilde{\phi}_{\Delta t H}(y^n), \\ \bar{y}^{n+1} &= X^{-1}(y^{n+1}). \end{aligned}$$

Thus we create a more accurate integrator

$$X^{-1} \circ \tilde{\phi}_{\Delta t H} \circ X.$$

If X is symplectic, there exists a modified Hamiltonian that can be used to assess the accuracy of the improved integrator. In practice it not necessary to pre- and post-process the numerical solution at every step. Rather the preprocessing $y^n = X(\bar{y}^n)$ need be done only at the beginning, for $n = 0$, and the post-processing $\bar{y}^{n+1} = X^{-1}(y^{n+1})$ only when output is desired. Moreover, because the transformation X is not part of the propagator, it can be approximated without any dire effects, and this approximation need not be symplectic. In the case of MD, the initial positions are not known precisely and the initial velocities are chosen at random, so preprocessing can be omitted. Similarly for the post-processing, if the accuracy of the computed quantities is unaffected by the high frequency error that is present in an unprocessed trajectory. Even in this extreme case where we forgo all pre- and post-processing, the concept is still a valuable theoretical or experimental tool [112] for comparing the intrinsic accuracy of two integrators.

It is natural to ask whether the order of accuracy of the leapfrog method itself might be increased. This is not, however, generally possible [69]. All that can be accomplished by a symplectic transformation X is to reduce the size of the coefficient of Δt^2 in an expansion of the error. For example, consider $\bar{y} = X^{-1}(y)$ defined by

$$\begin{aligned} \bar{q} &= q + \beta \Delta t^2 M^{-1} U_q(q), \\ \bar{p} &= \Big(I + \beta \Delta t^2 M^{-1} U_{qq}(q)\Big)^{-1} p. \end{aligned}$$

The accuracy of this processed leapfrog method can be assessed in terms of the modified Hamiltonian because the transformation X is symplectic. For leapfrog with the given processing, the modified Hamiltonian given by Eq. (6) becomes

$$\tilde{H}^{\text{proc}}(\bar{q}, \bar{p}) = H(\bar{q}, \bar{p}) + \Delta t^2 \left((\beta + \frac{1}{12}) \bar{p}^{\mathrm{T}} M^{-1} U_{qq}(\bar{q}) M^{-1} \bar{p} + (-\frac{1}{24} - \beta) U_q(\bar{q})^{\mathrm{T}} M^{-1} U_q(\bar{q}) \right). \quad (11)$$

The best choice of β is not obvious. However, the choice $\beta = -\frac{1}{16}$, which equalizes the coefficients, leads to energy conservation errors of only $O(\Delta t^4)$ for quadratic Hamiltonians. This is likely a near optimal choice for MD, where the dynamics is harmonic to a significant degree. A more complicated symplectic transformation which exactly conserves energy in the case of a quadratic Hamiltonian is given in [101]. A highly efficient approximation to symplectic post-processing transformations based on differencing is given in [70, 112]. If both the time integrator and the processing are symplectic, then the fluctuations in total energy of the processed solution represent a sampling of the perturbation to the Hamiltonian due to the finite time step Δt of the processed integrator. And with the removal of most of the non-growing part of the systematic error (which is rich in high frequency components), the energy of the processed trajectory becomes more revealing as an error indicator and hence as a detector of programming bugs and perhaps collision events. Indeed, this may be reason enough to actually perform the processing, or at least the post-processing.

3.6 Finite-precision effects

In reality, numerical methods are implemented in finite-precision arithmetic, and it is to be expected that rapid exponential growth of even tiny roundoff errors would overwhelm numerical trajectories well before the end of the simulation. Apparently these effects are not catastrophic; otherwise no one would be doing MD.

The best way to begin thinking about the matter is to realize that the effect of finite precision is to project the computed trajectory onto a finite lattice in phase space. Assuming that the trajectory is not terminated by overflow, it will eventually start repeating itself and settle into a limit cycle.

More can be said if we make certain idealized assumptions about the computation and if we assume that the lattice is uniform, which is the case for fixed point arithmetic but not for floating point. Building on the work of others, it is shown in [103] that the effect of computing the trajectory of a symplectic integrator on a uniform lattice is equivalent to applying that symplectic method to a slightly different Hamiltonian system with Hamiltonian $H_\varepsilon = H + O(\varepsilon^2)$ where ε is the machine epsilon. (The fact that the numerical flow is defined only for lattice points provides latitude in defining H_ε on the continuum.) Note that what is being said is that the finite precision trajectory is the exact *leapfrog* trajectory for a slightly different Hamiltonian, not the exact analytical trajectory for a different Hamiltonian. Nonetheless, this does mean that the finite precision map is symplectic.

In a subsequent paper, it is shown [108] that the idealized assumptions about the computation are unnecessary and that a *lattice leapfrog method* can be implemented in floating-point arithmetic at very little extra cost. Also shown is how the analysis extends to include errors other than those due to finite precision: the method of proof allows for errors in the force $F(q)$ that could be due to

- a fast electrostatic solver such as the fast multipole method,
- Lennard–Jones cutoffs, and
- errors due to unconverged iterations such as those needed to implement implicit integration schemes.

Being symplectic, however, is no guarantee of stability, and errors much larger than ε can produce instability [9].

Example 1 *The Kepler problem in polar coordinates is given by*

$$H(q,p) = \frac{p^2}{2m} - \frac{GMm}{q} + \frac{l^2}{2mq^2}.$$

One can choose units in q and t so that $H(q,p) = \frac{1}{2}p^2 - q^{-1} + \frac{1}{2}q^{-2}$. The potential energy function has a minimum at $q = 1$ and is negative for $\frac{1}{2} < q < +\infty$. Pieces of orbits for $\varepsilon = 2^{-13}, 2^{-14}, 2^{-15}, 2^{-16}$ are shown in Figs. 2–5. The time step is $\Delta t = 0.2$. We observe that increasing precision does not reduce the thickness of the orbit.

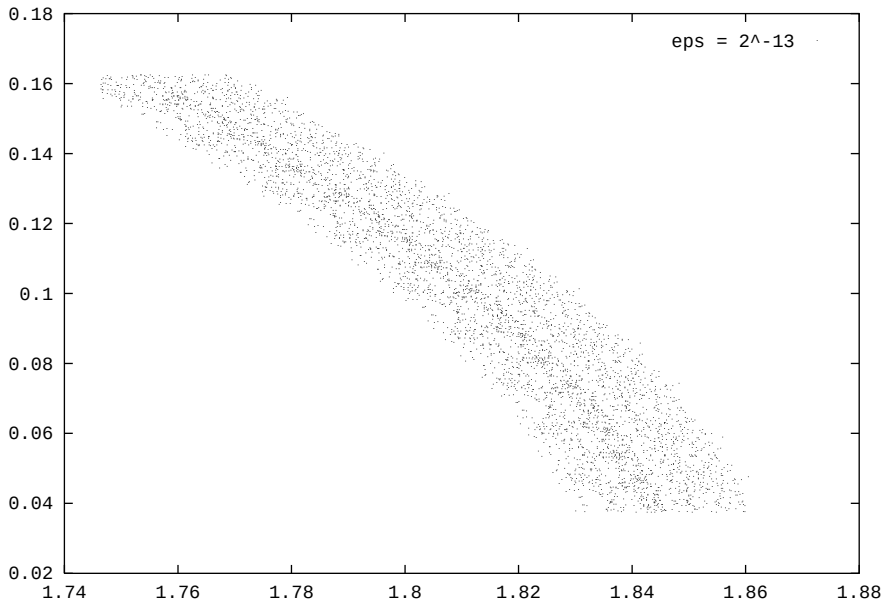

Figure 2: Piece of infinite orbit for polar-coordinate Kepler problem with $\varepsilon = 2^{-13}$.

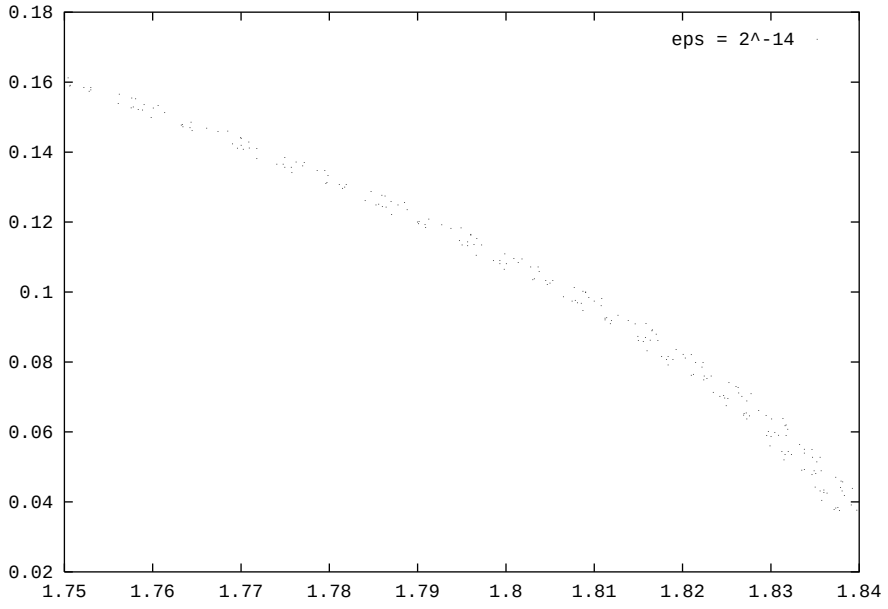

Figure 3: Piece of infinite orbit for polar-coordinate Kepler problem with $\varepsilon = 2^{-14}$.

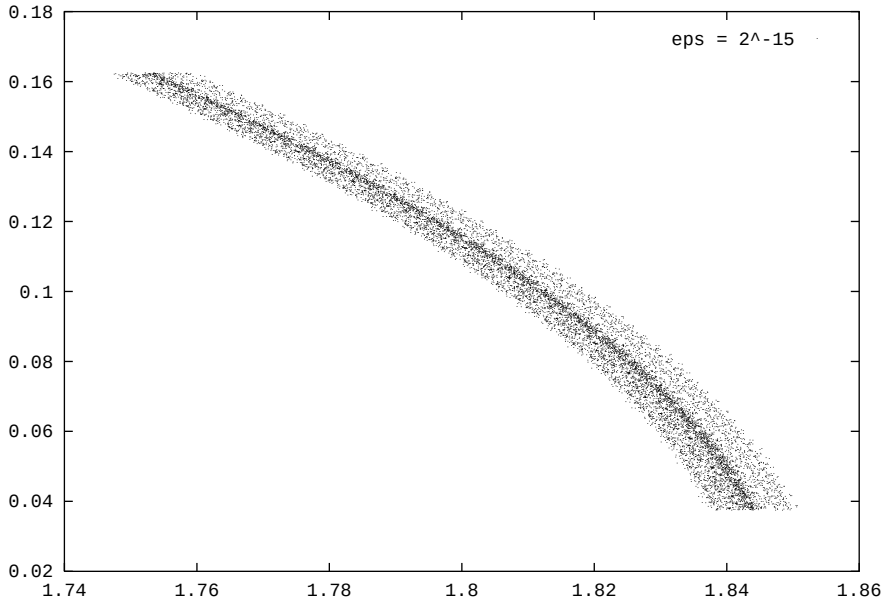

Figure 4: Piece of infinite orbit for polar-coordinate Kepler problem with $\varepsilon = 2^{-15}$.

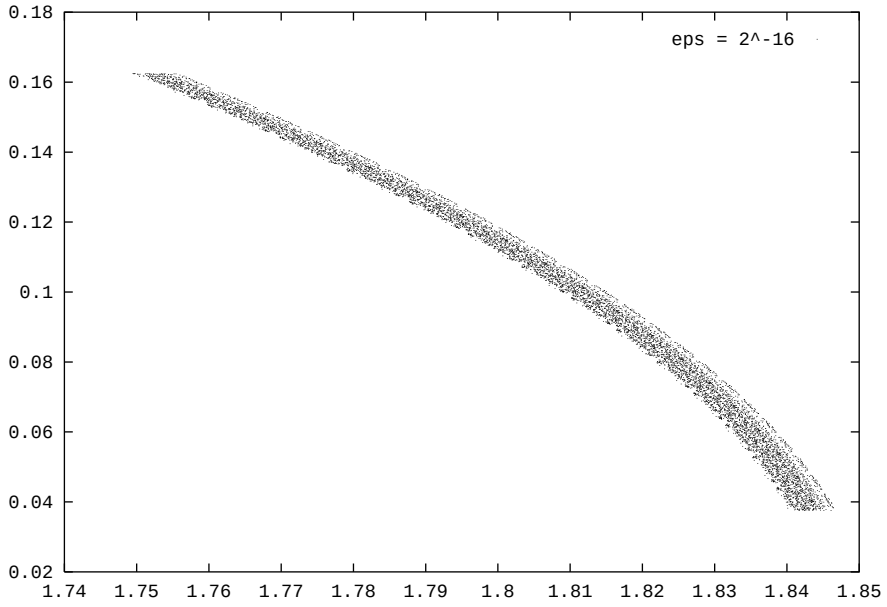

Figure 5: Piece of infinite orbit for polar-coordinate Kepler problem with $\varepsilon = 2^{-16}$.

4 Other Methods

Methods other than leapfrog have also been used extensively for MD. These other methods have mostly been of higher order and have all been non-symplectic. Notable examples are the "other" Gear methods (the higher order Störmer methods in Nordsieck form) [37, p. 154]. However, the leapfrog method has emerged as the favorite. Does the symplectic property of the leapfrog method explain its success in comparison with (non-symplectic) higher order methods? A case for symplectic integration is made at the end of this section.

Experience with numerical methods in general suggests that fourth-order methods are likely to be optimally efficient for low-accuracy solutions. For typical accuracies in MD there is some limited experimental evidence in favor of higher order symplectic integrators [88, 57].

Our brief look at other methods brings us into contact with a number of interesting results. One is the remarkable construction, by Yoshida, of symplectic methods of arbitrary order obtained by taking a basic low order symplectic method and executing a sequence of substeps of specially chosen sizes. Another is the Ge–Marsden result on the incompatibility between energy conservation and the symplectic property. A third is the result of Chawla on the optimal stability of the leapfrog among all explicit methods of standard type. Much more about numerical methods for Hamiltonian systems is to be found in the book by Sanz-Serna and Calvo [97].

4.1 A family of methods

We consider, as an example, a method more elaborate than leapfrog and potentially of higher order. Here is one step, given in algorithmic style:

$q = q + (\frac{1}{2} - a)\Delta t M^{-1} p;$
$p = p + (\frac{1}{2} - \frac{1}{2}a - \frac{1}{2}b)\Delta t F(q);$
$q = q + a\Delta t M^{-1} p;$
$p = p + (a + b)\Delta t F(q);$
$q = q + a\Delta t M^{-1} p;$
$p = p + (\frac{1}{2} - \frac{1}{2}a - \frac{1}{2}b)\Delta t F(q);$
$q = q + (\frac{1}{2} - a)\Delta t M^{-1} p;$

Here a and b are method parameters that can be chosen to optimize accuracy and stability. This particular parameterization is chosen for convenience in obtaining methods of order four and the method of Rowlands [95]. Note that the total increment for the positions q is exactly $\Delta t M^{-1} p$ and the total for the momenta is $\Delta t F(q)$. The choice $a = 0$ or $a + b = 1$ gives the midpoint form of the leapfrog method. This method is symplectic, as is any method based on splitting the Hamiltonian.

The modified Hamiltonian can be calculated using the BCH formula given in appendix A; the result is

$$\tilde{H}(q,p) = H(q,p) + \Delta t^2 \left(c_1 p^{\mathrm{T}} M^{-1} U_{qq} M^{-1} p + c_2 U_q^{\mathrm{T}} M^{-1} U_q \right)$$

where $c_1 = -\frac{1}{24} + \frac{1}{2}a^2 - \frac{1}{2}a^3 - \frac{1}{2}a^2 b$ and $c_2 = \frac{1}{12} - \frac{1}{4}a + \frac{1}{4}a^3 + \frac{1}{2}a^2 b + \frac{1}{4}ab^2$. With the post-processing transformation

$$\begin{aligned} \bar{q} &= q + \beta \Delta t^2 M^{-1} U_q(q), \\ \bar{p} &= \left(I + \beta \Delta t^2 M^{-1} U_{qq}(q) \right)^{-1} p, \end{aligned}$$

the modified Hamiltonian for the processed method is

$$\tilde{H}^{\mathrm{proc}}(\bar{q},\bar{p}) = H(\bar{q},\bar{p}) + \Delta t^2 \left((c_1 - \beta) \bar{p}^{\mathrm{T}} M^{-1} U_{qq}(\bar{q}) C M^{-1} \bar{p} + (c_2 + \beta) U_q(\bar{q})^{\mathrm{T}} M^{-1} U_q(\bar{q}) \right).$$

To attain effective order 4 requires $\beta = c_1 = -c_2$, which imposes the following condition on the two method parameters:

$$6ab^2 = 6a(1-a)^2 - 1.$$

This can be solved for b if either $a < 0$ or $a \geq \Omega$ where $\Omega = \frac{1}{3}(2 + 2^{1/3} + 2^{-1/3})$. Some interesting special choices are the following:

1. The choice $a = \frac{1}{2} - \frac{1}{2}\Omega$ and $b = \frac{1}{2} - \frac{3}{2}\Omega$ yields a method of *conventional* order 4 found first by Forest and Ruth [31, 16]. This scheme is simply the composition of three steps of the midpoint form of leapfrog with substeps of sizes $\Omega\Delta t$, $(1-2\Omega)\Delta t$, $\Omega\Delta t$. This scheme is, however, far from the best choice because of large error coefficients. There is one other scheme of conventional order 4, corresponding to choosing b to have the opposite sign. It has even larger error coefficients.

2. The choice $a = 1 - \Omega$ and $b = 2\Omega - 1$ yields a method with a maximal stability threshold for $\omega\Delta t$ for the harmonic oscillator $(\mathrm{d}^2/\mathrm{d}t^2)q = -\omega^2 q$ [70]. This can be regarded as the composition of a first order scheme followed by midpoint leapfrog followed by the "backward" form of the first order scheme with substeps of sizes $(\frac{1}{2} - \frac{1}{2}\Omega)\Delta t$, $\Omega\Delta t$, $(\frac{1}{2} - \frac{1}{2}\Omega)\Delta t$.

3. The limit $a \to 0$ and hence $b \to \infty$ yields a method found first by Rowlands [95, 71]:

   $$\begin{aligned} q &= q + \tfrac{1}{2}\Delta t M^{-1} p; \\ p &= p + \Delta t (I - \tfrac{1}{12}\Delta t^2 U_{qq}(q) M^{-1}) F(q); \\ q &= q + \tfrac{1}{2}\Delta t M^{-1} p; \end{aligned}$$

   It is not necessary to calculate the Hessian as such; rather one can form a Hessian–vector product in as little work as is needed to compute the force vector. In fact, the work is even less because expensive square roots and elementary function values can be re-used. (A more straightforward derivation of Rowlands' method is to choose $\beta = -\frac{1}{12}$ in (11) and then replace U in the leapfrog method by $U - \frac{1}{24}\Delta t^2 U_q^{\mathrm{T}}(q) M^{-1} U_q$ in order to cancel the dominant error term in the nearby Hamiltonian.)

4.2 Quest for accuracy and stability

We call the method above, in the case $a \neq 0$ and $a + b \neq 1$, a 3-stage method because it uses three force evaluations. Of course, even more stages can be used for the purpose of improving error coefficients, order of accuracy, and/or intervals of stability.

Awareness of the existence of symplectic numerical integrators is reported [17] to date back to unpublished work in 1956 by De Vogelaere [20]. The first published paper on symplectic methods is that of Ruth [96] in 1983, which shows that the leapfrog method possesses this property. Ruth also derives new symplectic methods, including an explicit method of order 3 not requiring analytical derivatives. Other early papers are [28, 17].

Then began a search for higher order methods based on splittings. In a rather remarkable development, Yoshida [130] gave a systematic construction for schemes of order $2r$ having 3^{r-1} stages. The first step of this construction creates the Forest–Ruth method from the leapfrog method. These methods all have negative fractional time steps and hence large coefficients and not-too-surprisingly large error coefficients. Therefore, current searches for higher order methods also seek modest error coefficients. However, for methods of order > 2 it is not possible to avoid negative fractional time steps in either q or p [105, 117, 38].

In general it is advantageous to use symmetric methods. If we have a good non-symmetric method, it can be composed with its adjoint (its "backward" form). The asymptotic expansion of the error for the composed method will have the same coefficients as the original method for even powers of Δt but the odd powers will be absent. And very often, but not always, the adjoint of a method is as efficient to execute as the original method.

A measure of (linear) stability is the length of the *interval of periodicity*, which is the longest interval of values of $\omega\Delta t$ starting at zero for which the integrator's propagation matrix has only eigenvalues of unit modulus and is power-bounded when applied to the test equation with Hamiltonian $\frac{1}{2}p^2 + \frac{1}{2}\omega^2 q^2$. A better measure is the length of the *scaled* interval of periodicity, based on the *scaled time step*, which is the time step Δt divided by the number of force evaluations. (A force evaluation at the very beginning of a step counts only as one half an evaluation if the same value is used also at the end of the previous step.) In this sense the leapfrog method is optimal—Chawla [18] has shown that no "conventional" explicit method has a scaled interval of periodicity which extends beyond $\omega\Delta t/m = 2$, where m is the number of force evaluations per step. The optimally stable effectively 4th order 3-stage method given above as item 2 is periodic for $(\omega\Delta t)/3 \leq 1.89\cdots$. Rowlands' method has a stability barrier of $\sqrt{3}$, if it is regarded as a 2-stage method. A long stability interval seems to have the added benefit of yielding smaller method and error coefficients; for example, the optimally stable scheme, just mentioned has an error coefficient that is smaller than that of the Forest–Ruth method by a factor of one hundred! Optimally stable schemes are extensively investigated in [80].

Multistep methods cannot be symplectic [46], but symmetric multistep methods [63] are reversible. These have been used by astronomers [90]. They exhibit a resonance due to parasitic principal roots of the difference equation.

4.3 The case for symplectic integration

A minimum reasonable requirement for an integrator for Hamiltonian systems is that it have an interval of periodicity of positive length, i.e., that it be neutrally stable for sufficiently small time steps Δt, when applied to a stable linear problem. This alone may not suffice (see [89] for an example of quadratic error growth where a symplectic method would give only linear error growth). Two stronger properties (neither of which implies the other) that might be adequate are the Liouville (or volume-preserving) property and the property of reversibility. A mapping $X(y)$ preserves volume if its Jacobian $\det X_y = 1$, which is a consequence of the property $X_y^{\mathrm{T}} J X_y = J$ of a symplectic mapping. Considered below are some consequences of these properties and their possible implication for molecular dynamics.

Perhaps the strongest motivation for using symplectic methods is given by *backward error analysis*. The numerical trajectories produced by any integrator can be expressed formally as the exact solution of perturbed system of ODEs whose right hand side function $f(y)$ is expressible as a formal series in powers of Δt. It has been shown [113, 97] that this perturbed system is a Hamiltonian system if and only if the integrator is symplectic. This series is generally divergent regardless of how small Δt is, although for linear problems it converges for small enough Δt. If the series is truncated at an appropriate place, the solution of the corresponding Hamiltonian system will differ only by an amount $O(\Delta t^p)$ from the numerical solution of the symplectic integrator on a time interval of length $O(1/\Delta t)$ where p is the order of accuracy. See [113, section IV3], [100], [44], [7], [47], and [91]. Unfortunately $1/\Delta t$ is much shorter than integration intervals in MD. Nonetheless, there is no evidence that the backward error analysis cannot be extended to much longer times.

As suggested from the discussion in Section 2.2 of the goals of MD simulations and the Liouville

equation, what we want from the numerical integrator are probability densities $\rho^n(y)$ that reasonably well approximate the theoretical densities $\rho(y, n\Delta t)$ even for very large $n\Delta t$. Hence, it would seem that the foremost concern is to have correct limiting behavior as $t \to \infty$. The numerical method ought to have a unique equilibrium density not too different from the correct one and it ought to maintain either the ergodic or mixing property so that convergence to equilibrium occurs. For the sake of discussion assume that the numerical integrator approximates the solution of a modified ODE system $(\mathrm{d}/\mathrm{d}t)y = \tilde{f}(y)$ on an infinite time interval, and also assume that there is an equilibrium probability density function $\tilde{\rho}(y)$ for ensembles of trajectories. Hence $\nabla^{\mathrm{T}}(\tilde{\rho}\tilde{f}) = 0$. Recall from sec. 2.2 that the true equilibrium probability density $\bar{\rho}(y)$ is a conserved quantity for the true dynamics. The density $\tilde{\rho}(y)$ will be conserved for the modified dynamics only if $\nabla^{\mathrm{T}}\tilde{f} = 0$. This apparently requires the integrator to have the Liouville property. Assume now that this is the case. There remains the concern that there is more than one equilibrium density for the modified ODE system or that the ergodic or mixing property is absent, preventing convergence to equilibrium. If this modified ODE system is Hamiltonian, then it is unlikely that there are additional integrals of motion because non-integrability is generic for Hamiltonian systems.

Concerning ergodicity, a comparison [17] with ordinary Runge–Kutta methods shows the symplectic method to have superior ergodic behavior. However, these Runge–Kutta methods do not satisfy the bare minimum linear stability requirement given at the beginning of this subsection, so this may be a case of the solution being attracted to a lower dimensional manifold by damping.

All known symplectic methods also conserve angular momentum [98, 131]. (Linear momentum is preserved by any standard method.) These seem like good properties to maintain if we want to have good infinite time behavior.

Stability is essential and KAM theory [4, p. 411] suggests superior stability properties for symplectic methods as a consequence of the preservation of invariant tori around an elliptic equilibrium. Also, stability for at least a time interval $O(1/\Delta t)$ is guaranteed by the conservation of energy for a nearby Hamiltonian.

Calculations requiring equilibrium MD can be performed rigorously with hybrid Monte Carlo methods which consist of a sequence of short MD trajectories, a small fraction of which are rejected. It has been shown that the property of detailed balance needed by this method is satisfied if the integrator is both reversible and volume-preserving [81].

A result of Ge and Marsden [35] shows that if a symplectic scheme conserves energy exactly for a Hamiltonian system having no integral of motion other than energy, then it computes the exact trajectory except for a reparameterization of time. This result suggests that small errors in energy conservation for a symplectic integrator correspond to a trajectory that in some sense also has small errors, and hence that the conservation of energy might be a useful error indicator *for symplectic methods* (whereas for non-symplectic methods it could be deceptively small). If the energy is left free to vary, it can also indicate the presence of programming bugs. For example a jump in energy is likely a bug. Also, a continual downward drift in energy is almost certainly a bug, if a volume-preserving integrator is being used. (There is less phase space volume at lower energy values.)

The use of a trajectory's history to enhance the accuracy of integration tends to be incompatible with being symplectic [87, 46]. As a consequence, the storage requirements of symplectic methods tend to be minimal.

Symplecticness does not imply reversibility. However, as suggested in the preceding subsection, symmetric methods, which are also reversible, offer more accuracy for the same computational cost. A case for time reversibility is given in [13].

5 Multiple Time Steps

Many of the terms that govern the dynamics of molecules correspond to distant Coulomb interactions which vary much more slowly than the short-range forces. Appropriate time steps vary from the 0.5 to 1 fs value needed to resolve the 9 fs period of the fastest normal modes associated with methyl groups to a value of perhaps 16 fs for the most distant electrostatics [32]. For bonded interactions a fixed time step is appropriate but for non-bonded interactions, the appropriate time step is distant-dependent and thus varies with time. This need—each interaction having an appropriate individual time step and non-bonded interactions having variable time steps—is met by multiple time stepping (MTS) methods, to a significant extent.

Multiple time stepping means the use of different time steps for different parts of the right-hand-side function $f(y)$ of a system of ODEs $(\mathrm{d}/\mathrm{d}t)y = f(y)$. In the case of an additive partitioning of $f(y)$, this is formally equivalent to a multirate method [36]. The use of MTS dates back to at least 1967 [50], in astronomy, and to 1978 in MD [115]. The program GROMOS [124] has long used a *twin-ranged method* in which longer-range forces are evaluated less frequently, for example, every ten steps. For the sake of discussion suppose that we separate the forces into two parts at time t^n,

$$F = F^{\text{fast}} + F^{\text{slow}},$$

and that the slow force is to be evaluated only once for the next two time steps. Commonly this is done by re-using $F^{\text{slow},n}$ at time t^{n+1}. Such a method is only first order accurate, and is not symplectic. In an effort to construct methods that retain the good behavior of the leapfrog/Verlet, a property termed "Verlet-equivalence" was proposed [40, 42]. This condition requires that the method reduce to the Verlet method in the case where all forces are zero except those associated with just (any) one of the time steps. The simplest such method, called Verlet-I, is based on delivering the slow force as widely spaced impulses:

$$M\frac{\mathrm{d}^2}{\mathrm{d}t^2}q = \sum_{n'}\frac{\Delta t}{m}\,\delta(t - n'\frac{\Delta t}{m})F^{\text{fast}}(q) + \sum_{n}\Delta t\,\delta(t - n\Delta t)F^{\text{slow}}(q).$$

Due to the possibility of resonance artifacts, this method was rejected in favor of a more elaborate scheme, Verlet-II, which distributes the slow force throughout the interval between slow force evaluations. Both methods retain the second order accuracy of Verlet. The latter method was tested using as many as 8 distance classes with sampling frequencies that grow by factors of two. A third method Verlet-X [8] is as accurate as Verlet-II [41, Tabelle 4.1, DC-1c] but is more economical in storage and is self-starting. The impulsive Verlet-I method was proposed independently and tested under the name r-RESPA [121], an equivalence first noted in [41, p. 28]. The Verlet-I method was observed to be symplectic in [8] and shown to be good at conserving energy unlike Verlet-II or Verlet-X. However, the possibility of resonance was confirmed on a simple nonlinear test problem. Nonetheless, the Verlet-I/r-RESPA/impulse MTS method gives acceptable results for time steps as great as 4 fs [56, 127] and has become quite popular.

The original technique for effecting variable time steps for non-bonded interactions was to move them from one time step class to another. A refinement of this technique that makes the transition smoother is used in [51, Verlet-IId]. An alternative method is the use by [121] of switching functions for partitioning. A similar technique based on artificial splittings of potentials is proposed in [109] as a method for doing symplectic variable time steps.

Methods based on natural or artificial splittings of Hamiltonians are quite application-specific. At the same time, conventional techniques for varying the time step leads to a loss of the symplectic property [110] and stability [107]. Recently, general techniques have been proposed for varying the

time step and maintaining reversibility [55] and even the symplectic property [45, 91]. These, of course, permit only a single time step to be used for all variables and interactions.

5.1 The Verlet-I/r-RESPA/impulse MTS method

To obtain a symplectic scheme, we partition the potential energy $U = U^{\text{fast}} + U^{\text{slow}}$ as a basis for partitioning the force vector. Then the integration for $\Delta t H$ is partitioned as

$(\Delta t/2)U^{\text{slow}}$,
$\Delta t(U^{\text{fast}} + K)$ numerically—m steps of leapfrog,
$(\Delta t/2)U^{\text{slow}}$.

To avoid complicated subscripting, we express the details algorithmically:

void *step*() {
    $p = p + \frac{1}{2}\Delta t F^{\text{slow}}(q)$;
    for ($k = 1$; $k \le m$; $k = k + 1$) {
        $p = p + \frac{1}{2}(\Delta t/m)F^{\text{fast}}(q)$;
        $q = q + (\Delta t/m)M^{-1}p$;
        $p = p + \frac{1}{2}(\Delta t/m)F^{\text{fast}}(q)$;
    }
    $p = p + \frac{1}{2}\Delta t F^{\text{slow}}(q)$;
}

Partitioning the potential energy among different interactions is adequate for bonded interactions of differing speeds but not for *nonbonded* interactions because their speed can vary greatly during the course of the simulation depending on the interparticle distance r. However, we can employ the machinery of MTS to vary the stepsize for each nonbonded interaction $U_{ij}(q)$ if we introduce an artificial splitting $U_{ij}(q) = U_{ij}^{\text{fast}}(q) + U_{ij}^{\text{slow}}(q)$ such that $U_{ij}^{\text{fast}}(q)$ vanishes for $\|\mathbf{r}_j - \mathbf{r}_i\|$ beyond some cutoff r_{cut} and $U_{ij}^{\text{slow}}(q)$ is slow for all q and never requires a small time step. The effect of this is to permit a large time step whenever $\|\mathbf{r}_j - \mathbf{r}_i\|$ exceeds the cutoff. An example of how to choose the splitting is given in the subsection that follows.

For a larger number of different time steps use a splitting

$$U = U_0 + U_1 + \cdots + U_\nu$$

where the term U_k is used with time step $2^{-k}\Delta t$. This defines a splitting of the force vector into terms $F_k(q) = -\nabla_q U_k(q)$. As an example, if $\nu = 2$, we can define the method from the following

progression of splittings of the Hamiltonian $\Delta t H$:

$$
\begin{array}{l}
\tfrac{1}{2}\Delta t U_0 \\
\left|\begin{array}{l}
+\ \tfrac{1}{4}\Delta t U_1 \\
\qquad \left|\begin{array}{l}
+\ \tfrac{1}{8}\Delta t U_2 \\
\qquad\qquad +\ \tfrac{1}{4}\Delta t K \\
+\ \tfrac{1}{8}\Delta t U_2
\end{array}\right| \\
\qquad \left|\begin{array}{l}
+\ \tfrac{1}{8}\Delta t U_2 \\
\qquad\qquad +\ \tfrac{1}{4}\Delta t K \\
+\ \tfrac{1}{8}\Delta t U_2
\end{array}\right| \\
+\ \tfrac{1}{4}\Delta t U_1
\end{array}\right| \\
\left|\begin{array}{l}
+\ \tfrac{1}{4}\Delta t U_1 \\
\qquad \left|\begin{array}{l}
+\ \tfrac{1}{8}\Delta t U_2 \\
\qquad\qquad +\ \tfrac{1}{4}\Delta t K \\
+\ \tfrac{1}{8}\Delta t U_2
\end{array}\right| \\
\qquad \left|\begin{array}{l}
+\ \tfrac{1}{8}\Delta t U_2 \\
\qquad\qquad +\ \tfrac{1}{4}\Delta t K \\
+\ \tfrac{1}{8}\Delta t U_2
\end{array}\right| \\
+\ \tfrac{1}{4}\Delta t U_1
\end{array}\right| \\
+\ \tfrac{1}{2}\Delta t U_0
\end{array}
$$

The potential energy terms can be merged into groups to produce the following set of *partial* force vectors:

$$
\begin{aligned}
F_0(q) + \tfrac{1}{2}F_1(q) + \tfrac{1}{4}F_2(q), \\
\tfrac{1}{4}F_2(q), \\
\tfrac{1}{2}F_1(q) + \tfrac{1}{4}F_2(q), \\
\tfrac{1}{4}F_2(q), \\
F_0(q) + \tfrac{1}{2}F_1(q) + \tfrac{1}{4}F_2(q).
\end{aligned}
$$

Below is given an algorithm for general ν, which loops through 2^ν substeps in each step. Before the very first step a *partial* force vector F is computed by

$$F = F_0(q) + 2^{-1}F_1(q) + \cdots + 2^{-\nu}F_\nu(q);$$

Then each step proceeds as follows:

void *step*() {
  for ($i = 1$; $i \le 2^\nu$; $i = i + 1$) {
    $p = p + \frac{1}{2}\Delta t F$;
    $q = q + 2^{-\nu}\Delta t M^{-1} p$;
    l = the smallest integer such that $2^{\nu-l}$ divides i;
    $F = 2^{-l}F_l(q) + 2^{-l-1}F_{l+1}(q) + \cdots + 2^{-\nu}F_\nu(q)$;
    $p = p + \frac{1}{2}\Delta t F$;
  }
}

If the algorithm is implemented as described above, it will not be much faster than a single time step algorithm for three reasons:

1. At time steps when slower forces are being evaluated, there are more interactions to be evaluated, e.g., $F_0(q) + \frac{1}{2}F_1(q) + \frac{1}{4}F_2(q)$ instead of just $F(q)$.

2. Even though the evaluations of most interactions are avoided at most time steps because they are zero, there remains the cost of determining whether or not to evaluate an interaction. This test can be done by computing the square of a distance and comparing this to the square of the cutoff, thus avoiding a square root calculation. However, this alone will not provide dramatic savings in computing time.

3. If the shortest possible time step $2^{-\nu}\Delta t$ is chosen smaller than what is needed for bonded forces in order to resolve rare close encounters between a non-bonded pair of atoms, it would inefficient to descend to this level except when necessary.

Remedies for these inefficiencies are given in [49], two of which are described in section 5.3. It is helpful first to discuss in greater detail the method of partitioning the non-bonded interactions.

5.2 Partitioning of interactions

There are two issues: (i) how to assign appropriate time steps to each interaction, and (ii) how to split a non-bonded interaction. The choice of parameters in the splitting technique depends on the assignment of time steps.

A rule for the assignment of time steps to differing interactions in a multiple time stepping scheme is given in [69]. It is based on the spectral radius of the mass-weighted Hessian of an interaction potential $U(q)$:

$$\Delta t \propto \left(\max_q \rho(M^{-1/2} U_{qq}(q) M^{-1/2}) \right)^{-1/2}.$$

This rule is derived for bonded interactions from a truncation error analysis. It is a reasonable rule of thumb for non-bonded interactions as well.

The question of how to split a non-bonded interaction is addressed first for electrostatics. The potential energy term is given by

$$U_{ij}(q) = \frac{1}{2} C_{ij} \phi(\|\mathbf{r}_j - \mathbf{r}_i\|^2)$$

where

$$\phi(s) = 2s^{-1/2}.$$

Note that ϕ is defined in such a way that its argument $\|\mathbf{r}_j - \mathbf{r}_i\|^2$ is a smooth (in fact, quadratic) function of q. The appropriate time step depends on the value of s. Let L denote the following linear operator

$$L\gamma(s) = \begin{cases} \gamma(r_{\text{cut}}^2) + (s - r_{\text{cut}}^2)\gamma'(r_{\text{cut}}^2), & s \le r_{\text{cut}}^2, \\ \gamma(s), & s \ge r_{\text{cut}}^2, \end{cases}$$

in which a function $\gamma(s)$ is replaced by its linear Taylor interpolant for $s \le r_{\text{cut}}^2$. Define the splitting

$$\phi^{\text{slow}}(s) = L\phi(s), \quad \phi^{\text{fast}}(s) = (1 - L)\phi(s).$$

Then $U_{ij}(q) = \frac{1}{2} C_{ij} \phi^{\text{slow}}(\|\mathbf{r}_j - \mathbf{r}_i\|^2) + \frac{1}{2} C_{ij} \phi^{\text{fast}}(\|\mathbf{r}_j - \mathbf{r}_i\|^2)$ with

$$\frac{1}{2}\phi^{\text{slow}}(r^2) = \begin{cases} \frac{3}{2} r_{\text{cut}}^{-1} - \frac{1}{2} r_{\text{cut}}^{-3} r^2, & r \le r_{\text{cut}}, \\ r^{-1}, & r \ge r_{\text{cut}}, \end{cases} \qquad \frac{1}{2}\phi^{\text{fast}}(r^2) = \begin{cases} r^{-1} - \frac{3}{2} r_{\text{cut}}^{-1} + \frac{1}{2} r_{\text{cut}}^{-3} r^2, & r \le r_{\text{cut}}, \\ 0, & r \ge r_{\text{cut}}. \end{cases}$$

Each piece has a continuous first derivative. In such splittings it is desirable to avoid cancellation to keep the temporal discretization error small [116].

The technique applies also to the Lennard–Jones potential, given by

$$U_{ij}^{\mathrm{LJ}}(q) = \frac{1}{2}\varepsilon_{ij}\phi^{\mathrm{LJ}}(\|\mathbf{r}_j - \mathbf{r}_i\|^2/\sigma_{ij}^2) - \varepsilon_{ij}$$

where

$$\phi^{\mathrm{LJ}}(s) = 2(2s^{-3} - 1)^2.$$

If r_{cut} is chosen where $\phi^{\mathrm{LJ}}(r^2/\sigma_{ij}^2)$ takes its minimum value, the result is a splitting of the Lennard–Jones potential into a repulsive part and an attractive part, known as the Weeks–Chandler–Andersen (WCA) breakup .

We can do multiple time stepping for an electrostatic potential energy term by defining a multiple splitting of the function $\phi(s) = 2s^{-1/2}$. We partition the range of values of s by means of a sequence of cutoffs $r_1 > r_2 > \cdots > r_k > \cdots > 0$ such that $2^{-k}\Delta t$ is the appropriate time step when $r_{k+1}^2 \le s < r_k^2$. Here it is convenient to define $r_0 = +\infty$. Then we construct an *artificial* splitting

$$\phi = \phi_0 + \phi_1 + \cdots + \phi_k + \cdots \tag{12}$$

such that

1. $2^{-k}\Delta t$ is an appropriate time step for ϕ_k and
2. $\phi_k(s) = 0$ for $s \ge r_k^2$.

This means that if $r_{k+1}^2 \le s < r_k^2$, then $2^{-k}\Delta t$ is the smallest time step needed for any nonzero term in the splitting. To obtain the splitting, we can repeatedly apply the two-time-step idea

$$\begin{aligned}
\phi(s) &= \phi_0(s) + \psi_1(s) \\
&= \phi_0(s) + \phi_1(s) + \psi_2(s) \\
&= \phi_0(s) + \phi_1(s) + \phi_2(s) + \psi_3(s) \\
&\vdots \quad .
\end{aligned}$$

This can be done by the following recursive construction:

$$\psi_0(s) = \phi(s), \tag{13}$$

and, for $k = 1, 2, \ldots, \nu$,

$$\phi_{k-1}(s) = L_k\psi_{k-1}(s), \quad \psi_k(s) = (1 - L_k)\psi_{k-1}(s) \tag{14}$$

where L_k denotes the linear operator

$$L_k\gamma(s) = \begin{cases} \gamma(r_k^2) + (s - r_k^2)\gamma'(r_k^2), & s \le r_k^2, \\ \gamma(s), & s \ge r_k^2. \end{cases}$$

In particular,

$$\frac{1}{2}L_k\phi(r^2) = \begin{cases} \frac{3}{2}r_k^{-1} - \frac{1}{2}r_k^{-3}r^2, & r \le r_k, \\ r^{-1}, & r \ge r_k. \end{cases}$$

The recurrence (13), (14) can be solved by using the fact that $L_kL_{k-1} = L_{k-1}$ (linear Taylor interpolation reproduces a linear polynomial) to get

$$\phi_{k-1}(s) = (L_k - L_{k-1})\phi(s), \quad \psi_k(s) = (1 - L_k)\phi(s).$$

The second of these holds also for $k = 0$ if we define $L_0 = 0$.

5.3 Efficient implementation

Given here are remedies for the two of the three inefficiencies listed at the end of section 5.1. This is also a good place to point out that the force F_0 normally consists only of long-range electrostatics and that this is most efficiently done by some fast tree algorithm or grid-based solver.

We consider first how to avoid the extra work created by the different parts of a split force. Consider the calculation of the lth level partial forces

$$F = 2^{-l}F_l(q) + 2^{-l-1}F_{l+1}(q) + \cdots + 2^{-\nu}F_\nu(q);$$

With U_k defined according to eq. (12) the contribution to the partial force vector of the electrostatic force on particle i due to particle j is the gradient of

$$\frac{1}{2}C_{ij}\chi_l(\|\mathbf{r}_j - \mathbf{r}_i\|^2),$$

where

$$\chi_l(s) = 2^{-l}\phi_l(s) + 2^{-l-1}\phi_{l+1}(s) + \cdots + 2^{-\nu+1}\phi_{\nu-1}(s) + 2^{-\nu}\psi_\nu(s).$$

This simplifies:

$$\begin{aligned}\chi_l(s) &= (2^{-l}(L_{l+1} - L_l) + 2^{-l-1}(L_{l+2} - L_{l+1}) + \cdots + 2^{-\nu+1}(L_\nu - L_{\nu-1}) + 2^{-\nu}(1 - L_\nu))\phi(s) \\ &= (-2^{-l}L_l + 2^{-l-1}L_{l+1} + 2^{-l-2}L_{l+2} + \cdots + 2^{-\nu}L_\nu + 2^{-\nu})\phi(s).\end{aligned}$$

Let $k(s)$ be defined by

$$r^2_{k(s)+1} \le s < r^2_{k(s)}$$

unless $s < r^2_\nu$, in which case $k(s)$ is defined to be ν. Hence, $L_k\phi(s) = \phi(s)$ for $k \ge k(s) + 1$. There are two cases. If $k(s) < l$, then *by construction*

$$\chi_l(s) = 0.$$

If $k(s) \ge l$,

$$\begin{aligned}\chi_l(s) &= (-2^{-l}L_l + 2^{-l-1}L_{l+1} + 2^{-l-2}L_{l+2} + \cdots + 2^{-k(s)}L_{k(s)})\phi(s) + 2^{-k(s)}\phi(s) \\ &= a_{l,k(s)}s + b_{l,k(s)} + 2^{-k(s)}\phi(s)\end{aligned}$$

where $a_{l,k(s)}$ and $b_{l,k(s)}$ are constants depending on $\phi(s)$ and the cutoff radii. The values $b_{l,k(s)}$ are not needed for force calculations and the values $a_{l,k(s)}$ and $2^{-k(s)}$ can be pre-computed and stored in tables.

Considered here is the question of avoiding unnecessary distance calculations. As noted already, the test $\|\mathbf{r}_j - \mathbf{r}_i\|^2 < r^2_{\text{cut}}$ is considerably cheaper to perform than is the computation of a force with potential energy $C_{ij}\|\mathbf{r}_j - \mathbf{r}_i\|^{-1}$, because of the additional square root calculation. Nonetheless, it is highly desirable to avoid doing this test in the vast majority of cases. The cost of doing each collection of fast force evaluations should be only linear in the number of particles N (although cubic in the cutoff r_k). A linear time algorithm can be attained by using "cell lists" [26]. The computational box is divided into cubic or nearly cubic cells, and periodically, e.g., at the beginning of every (full) step, a list is constructed for each cell of the atoms that it contains. (For efficiency the looping is done over atoms.) Also, we choose a *safety margin* δ to be a distance that is unlikely to be traversed by any atom within the chosen period. This is monitored by checking that each particle stays within a distance δ of its container during the period. In this way distance testing is needed only for those atoms whose containers are within a distance $r_k + 2\delta$ of each other. Efficiency may be further enhanced by using the cell lists to construct a neighbor list, which would contain a list of all atom pairs (i, j) for which $r_{ij} \le r_{\text{cut}} + 2\delta$.

5.4 Mollified impulse MTS methods

Stability analysis of the Verlet-I/r-RESPA/impulse MTS method [34] shows that for linear problems it is numerically unstable not only for time steps near the period of the fastest normal mode but also for time steps just less than half the period. Also an "order reduction" effect is observed; i.e., the second order accuracy in terms of the *long* time steps is not independent of the fast force. The stability limitations are observed in practice for MD simulations with time steps just less than the half period of the fastest normal mode, which is about 5 fs [56, 9]. The resonance observable in various graphs in [59] is almost certainly the same phenomenon.

The "five femtosecond time step barrier" of the impulse method is overcome by the *mollified* impulse method (MOLLY) [34], which replaces $U^{\text{slow}}(q)$ by $U^{\text{slow}}(\mathcal{A}(q))$ where $\mathcal{A}(q)$ is a time averaging of vibrational motion due to the fastest forces, e.g., bond stretching and angle bending. This modification has the effect of replacing $F^{\text{slow}}(q)$ by $\mathcal{A}_q(q)^{\text{T}} F^{\text{slow}}(\mathcal{A}(q))$. The main benefit is to filter out those components of the impulse that would excite the fastest forces. Stability can be shown for a linear problem with two degrees of freedom for time steps less than a period, as can uniform second order accuracy [34]. Various averaging functions $\mathcal{A}(q)$ with B-spline weights are tested for MD in [111].

More effective than averaging is a choice for $\mathcal{A}(q)$ based on projection [58]. If $U^{\text{fastest}}(q)$ consists only of bonded terms, we can write down a set of equations

$$g^k(q) = 0, \quad k = 1, 2, \ldots, \mu,$$

that represent equilibrium, $U_q^{\text{fastest}}(q) = 0$, for the fastest interactions. These are, in fact, distance constraints, and we can write

$$g^k(q) = \|\mathbf{r}_{j(k)} - \mathbf{r}_{i(k)}\|^2 - l_k^2.$$

Projection onto these constraints can be accomplished by defining

$$\mathcal{A}(q) = q + M^{-1} \sum_{l=1}^{\mu} \lambda_l g_q^l(q)$$

where the $\{\lambda_l\}$ are chosen to satisfy $g(\mathcal{A}(q)) = 0$. The solution of these quadratic systems of equations is a topic of the section that follows. Experiments indicate an increase of about 50% in the longest time step for this "equilibrium" version of MOLLY over the impulse method for a given level of energy conservation. They also indicate that the time step is being limited by the fastest normal modes.

6 Constrained Dynamics

Even with multiple time stepping the time step is restricted by the small-amplitude high-frequency components of the motion that are due to bond-length stretching. To enable larger time steps, it is common to make some or all of the bond lengths rigid. Sometimes angles are also constrained. This is accomplished by appending constraint equations to the equations for Newton's law of motion and by including one Lagrange multiplier in the force term for each constraint equation. This results in a system of differential equations with quadratic constraints. As in the previous section we express these constraints as

$$g(q) = 0 \qquad \text{where} \quad g^k(q) = \|\mathbf{r}_{j(k)} - \mathbf{r}_{i(k)}\|^2 - l_k^2, \quad k = 1, 2, \ldots, \mu.$$

Constraints on angles can be imposed as distance constraints if bond lengths are also being constrained. The set of constraints is used to determine Lagrange multipliers in the equations of motion:

$$\frac{\mathrm{d}}{\mathrm{d}t}q = M^{-1}p, \quad \frac{\mathrm{d}}{\mathrm{d}t}p = F(q) + g_q(q)^{\mathrm{T}}\lambda.$$

Constraints on positions clearly reduce the number of degrees of freedom in the system. Differentiating the position constraints gives velocity constraints:

$$g_q(q)M^{-1}p = 0.$$

(Differentiating the velocity constraints gives a linearly implicit formula for the Lagrange multipliers, so this is an index 3 differential-algebraic equation [11].) Hence phase space is of dimension $6N - 2\mu$. Ideally, the initial conditions satisfy the constraints. The differential system is Hamiltonian in an appropriate set of $6N - 2\mu$ coordinates [66].

The most important question concerns the inaccuracy introduced by imposing rigidity. Computer simulations indicate that the application of constraints to bond lengths has only a slight effect on dynamics, but that constraining all bond lengths and angles—*torsion dynamics*—has a significant effect. The argument justifying the use of constraints is based on averaging. In the case of the canonical (NVT) ensemble this averaging introduces a correction to the potential, first obtained by Fixman [30]. This correction is small for bond length constraints but not for angle constraints. The Fixman potential for angle constraints is generally difficult to compute, but a practical approximation has been found [94, 93], having the form of a correction to the dihedral potential. Rigid angles also introduce excessive rigidity in the structure, which is often compensated for by softening the Lennard–Jones potential. A method termed *soft-SHAKE* [94] avoids this artificial rigidity by formulating the constrained form of the ODEs as a singular perturbation expansion and taking the expansion a little further.

Because of a small mass, interactions involving hydrogen are faster than similar interactions involving heavy atoms, and it makes sense to select only such interactions for constraints. In particular, it is common to do simulations with constraints on only the lengths of covalent bonds to hydrogens. Also, independently of whether or not other constraints are imposed, it is common to use fully rigid water models.

6.1 Discretization

A popular generalization of the leapfrog method to constrained dynamics is the SHAKE method [122], which replaces the drift step of leapfrog by a "targeted" drift. The Lagrange multipliers at one time level n are used to satisfy the constraint equations at time level $n + 1$. An enhancement known as RATTLE [3] also forces velocities to satisfy their constraints. However, these adjustments to the velocities are effectively undone by the next SHAKE step, and the dynamics is unaffected. Hence, the RATTLE step is just cosmetic. It has been shown that RATTLE, and therefore SHAKE, is reversible and symplectic [67].

Given below is the algorithm for SHAKE with the RATTLE option. The step begins with values q^n, p^n, F^n obtained from the previous step. Then we compute

half-kick:

$$p^{n+\epsilon} = p^n + \frac{1}{2}\Delta t F^n,$$

SHAKE:

$$p^{n+1-\epsilon} \quad = \quad p^{n+\epsilon} + \Delta t g_q(q^n)^{\mathrm{T}}\lambda^n,$$

$$
\begin{aligned}
q^{n+1} &= q^n + \Delta t M^{-1} p^{n+1-\epsilon} \\
&\quad \text{where} \;\; \lambda^n \;\; \text{satisfies} \;\; g(q^{n+1}) = 0,
\end{aligned}
$$

half-kick:

$$
\begin{aligned}
F^{n+1} &= F(q^{n+1}), \\
\bar{p}^{n+1} &= p^{n+1-\epsilon} + \frac{1}{2}\Delta t F^{n+1},
\end{aligned}
$$

$*$RATTLE:

$$
\begin{aligned}
p^{n+1} &= \bar{p}^{n+1} + \Delta t g_q(q^{n+1})^{\mathrm{T}} \lambda_{\mathrm{p}}^{n+1} \\
&\quad \text{where} \;\; \lambda_{\mathrm{p}}^{n+1} \;\; \text{satisfies} \;\; g_q(q^{n+1}) M^{-1} p^{n+1} = 0.
\end{aligned}
$$

The SHAKE step involves solving nonlinear equations and this raises the question of existence and uniqueness of the solution. Even with one constraint the quadratic will have two solutions and in extreme cases none. More generally the kth constraint equation is dominated by the kth unknown Lagrange multiplier and is quadratic in this unknown. Hence, there are typically about 2^μ solutions. It is suggested that the physically correct solution is one obtained by analytical continuation as the atoms drift from q^n to q^{n+1}. If analytical continuation is terminated by bifurcation, then the time step is simply too large.

Application of constraints to a small molecule frequently results in a completely rigid molecule, in which case it is feasible to use a more direct description of rigid body dynamics. If $\mathbf{r}_1$, $\mathbf{r}_2$, ..., $\mathbf{r}_m$ are the atomic positions of a rigid molecule, they can be expressed as

$$
\mathbf{r}_i = \mathbf{c} + Q\mathbf{d}_i
$$

where $\mathbf{c}$ is a varying center of mass, Q is a varying orthogonal rotation matrix, and the $\mathbf{d}_i$ are constant positions of the atoms in body-fixed coordinates. The rotation matrix Q embraces three degrees of freedom, but in order to avoid singularities it has been common to use quaternions [25], which involves four variables with one constraint. The equations of motion for quaternions can be expressed as a constrained Hamiltonian system. However, positions and momenta are coupled in the expression for the kinetic energy, and this forces the use of implicit symplectic methods.

A better foundation for discretization uses as variables all nine elements of the rotation matrix Q (see [76, 92]) subject to the six constraints $g(Q) = Q^{\mathrm{T}}Q - I = 0$ (usually enforced by the incorporation of Lagrange multipliers). Again, this choice of variables admits a constrained Hamiltonian formulation, which can be solved using the symplectic SHAKE algorithm, as discussed in [76]. Recently, a variation of this method was evaluated for molecular applications in [62], with an efficient treatment of the nonlinear equations for the Lagrange multipliers, resulting in improved conservation of energy for simulations of a dipolar soft-sphere fluid compared to a more traditional quaternion-based scheme. Moreover, explicit symplectic integration is possible. The idea is to split the rotational Hamiltonian into kinetic and potential terms, and integrate the kinetic term (describing a free rigid body) by reduction to the Euler equations for the free rigid body [77, 92]. This method was recently refined and applied to treat rigid body molecular problems in [23], with manifest improvements in observed long term stability and accuracy.

6.2 Solution of the nonlinear equations

There remains the question of solving the quadratic system of constraint equations. If we define $\bar{q} = q^n + \Delta t M^{-1} p^{n+\epsilon}$ and $d^l = \Delta t^2 M^{-1} g_q^l(q^n)$, then the constraint equations simplify somewhat to

$$
q^{n+1} = \bar{q} + \sum_l \lambda_l d^l
$$

where the $\{\lambda_l\}$ solve

$$\{g^k(q^{n+1}) = 0\}.$$

Nonlinear Gauss–Seidel applied to these equations for the $\{\lambda_l\}$ is

$\lambda_k = 0$, $k = 1, 2, \ldots, m$;
while (not converged)
  for ($k = 1$; $k \le m$; $k = k + 1$)
    solve $g^k(\bar{q} + \sum_l \lambda_l d^l) == 0$ for λ_k;

Each equation is quadratic in its associated unknown. A further simplification is to replace the quadratic solve by a single Newton–Raphson iteration:

$$\text{solve } \lambda_k = \lambda_k - \frac{g^k(\bar{q} + \sum_l \lambda_l d^l)}{g^k_q(\bar{q} + \sum_l \lambda_l d^l)^{\mathrm{T}} d^k};$$

yielding Newton–Gauss–Seidel. This algorithm is the originally proposed and popular way of solving the SHAKE constraint equations. The number of iterations can be halved by using over-relaxation [5] (Newton-SOR). The same study finds that the use of full Newton methods (with matrices) is no faster than Newton-SOR in CPU time.

7 Constant-Temperature and Constant-Pressure Ensembles

For equilibrium MD calculations it is preferable to sample from ensembles more similar to physiological conditions of constant temperature and pressure (NpT). Methods for doing this are given in [82, 75, 74, 119, 73, 27].

As an illustration of the ideas, we consider two ensembles for which the methods are simpler: a constant-temperature ensemble (NVT) and a constant-pressure ensemble (isobaric–isoenthalpic). The methods considered are based on extending the ODE system with an additional degree of freedom [1, 65]. The extended system is designed to yield time averages that are the same as those for the desired ensemble. Both extended systems have formulations that are Hamiltonian but non-separable. Nonetheless, both can be integrated by symplectic extensions of leapfrog that are effectively explicit.

7.1 Constant-temperature ensembles

The instantaneous temperature of an unconstrained system at time t is given by

$$T(t) = \frac{2}{dk_{\mathrm{B}}} \sum_i \frac{1}{2} m_i |\mathbf{v}_i|^2$$

where k_{B} is Boltzmann's constant and d is the number of degrees of freedom ($3N - 3$ if the total momentum is fixed). The temperature T is the time average of $T(t)$.

The extended system method was introduced by Nosé [85] with additional work by Hoover [54]. The idea is to model the dynamics of a system in contact with a thermal reservoir, so an additional degree of freedom is included, which represents the reservoir, and the extended system is simulated. Heat flows dynamically between the reservoir and the system, so the reservoir has a "thermal inertia" associated with it analogous to the mass of a particle.

Hoover's equations of motion for the system are

$$\frac{\mathrm{d}}{\mathrm{d}t} q = M^{-1} p, \quad \frac{\mathrm{d}}{\mathrm{d}t} p = F(q) - \xi p,$$

where the friction coefficient ξ is an extra degree of freedom, which satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\xi = \frac{1}{Q}\left(p^{\mathrm{T}}M^{-1}p - dk_{\mathrm{B}}T\right)$$

where Q is a thermal inertia parameter. It can be seen that ξ changes so that it restores the kinetic energy to its desired value.

Hoover's equations are not a Hamiltonian system, but they can be formulated as such, which is how Nosé originally expressed them. Introduce an integrating factor s, $(\mathrm{d}/\mathrm{d}t)s = \xi s$, for the momentum equation and a fictitious momentum

$$\bar{p} = sp,$$

and the system becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}q = s^{-1}M^{-1}\bar{p}, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\bar{p} = sF(q), \tag{15}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}s = \xi s, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\xi = \frac{1}{Q}\left(s^{-2}\bar{p}^{\mathrm{T}}M^{-1}\bar{p} - dk_{\mathrm{B}}T\right). \tag{16}$$

Define $p_{\mathrm{s}} = Q\xi$ and we have nearly a Hamiltonian system,

$$\frac{1}{s}\frac{\mathrm{d}}{\mathrm{d}t}q = s^{-2}M^{-1}\bar{p}, \qquad \frac{1}{s}\frac{\mathrm{d}}{\mathrm{d}t}\bar{p} = F(q), \tag{17}$$

$$\frac{1}{s}\frac{\mathrm{d}}{\mathrm{d}t}s = \frac{1}{Q}p_{\mathrm{s}}, \qquad \frac{1}{s}\frac{\mathrm{d}}{\mathrm{d}t}p_{\mathrm{s}} = s^{-3}\bar{p}^{\mathrm{T}}M^{-1}\bar{p} - dk_{\mathrm{B}}Ts^{-1}, \tag{18}$$

for Hamiltonian

$$\frac{1}{2}s^{-2}\bar{p}^{\mathrm{T}}M^{-1}\bar{p} + U(q) + \frac{1}{2Q}p_{\mathrm{s}}^2 + dk_{\mathrm{B}}T\ln s.$$

All that remains is to reparameterize time according to

$$\frac{\mathrm{d}}{\mathrm{d}\tau}t = \frac{1}{s},$$

where τ is fictitious time and the system becomes Hamiltonian:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}q = s^{-2}M^{-1}\bar{p}, \qquad \frac{\mathrm{d}}{\mathrm{d}\tau}\bar{p} = F(q), \tag{19}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}s = \frac{1}{Q}p_{\mathrm{s}}, \qquad \frac{\mathrm{d}}{\mathrm{d}\tau}p_{\mathrm{s}} = s^{-3}\bar{p}^{\mathrm{T}}M^{-1}\bar{p} - dk_{\mathrm{B}}Ts^{-1}. \tag{20}$$

The extended Hamiltonian is the sum of the usual kinetic and potential energy plus an extra kinetic energy term $p_{\mathrm{s}}^2/(2Q)$ and an extra potential energy term $dk_{\mathrm{B}}T\ln s$.

Although the system is non-separable, it has enough special structure to permit the use of explicit symplectic integrators. For example, the leapfrog method for general Hamiltonians $H(q,p)$,

$$\begin{aligned}
p^{n+1/2} &= p^n - \Delta t H_q(q^n, p^{n+1/2}),\\
q^{n+1/2} &= q^n + \Delta t H_p(q^n, p^{n+1/2}),\\
q^{n+1} &= q^{n+1/2} + \Delta t H_p(q^{n+1}, p^{n+1/2}),\\
p^{n+1} &= p^{n+1/2} - \Delta t H_q(q^{n+1}, p^{n+1/2}),
\end{aligned}$$

is explicit if in the first equation $\bar{p}$ is updated before p_{s} and in the third equation s is updated before q. Alternatively, one can advance by analytical integrations for Hamiltonian $\frac{1}{2}\Delta t(K + \frac{1}{2}Q^{-1}p_{\mathrm{s}}^2)$ followed by $\Delta t(U + dk_{\mathrm{B}}T \ln s)$ followed by $\frac{1}{2}\Delta t(K + \frac{1}{2}Q^{-1}p_{\mathrm{s}}^2)$.

Canonical ensemble (NVT) averages can be computed from time averages using real momentum and real time, under the assumption that the extended system is ergodic. More precisely, it can be shown that

$$\lim_{t\to\infty} \frac{\int_0^t g(q(\tau), s(\tau)^{-1}\bar{p}(\tau))s(\tau)^{-1}\,\mathrm{d}\tau}{\int_0^t s(\tau)^{-1}\,\mathrm{d}\tau} = \frac{\int\int g(q,p)\exp(-H(q,p)/(k_{\mathrm{B}}T))\mathrm{d}q\mathrm{d}p}{\int\int \exp(-H(q,p)/(k_{\mathrm{B}}T))\mathrm{d}q\mathrm{d}p}.$$

Numerical integration of the extended system yields data at unequally spaced points in real time. This is awkward for computing time correlation functions. However, there is a real time Hamiltonian formulation of Nosé–Hoover dynamics, termed the Nosé–Poincaré method [10]. It is based on a Poincaré transformation of the Nosé extended Hamiltonian with thermostat and yields canonical ensemble time averages. An efficient symplectic integrator is provided. The paper includes substantial numerical experiments as well as many details such as constraints, rigid bodies, and "Nosé chains."

7.2 Constant-pressure ensembles

The instantaneous pressure is

$$P(t) = \frac{1}{3V}\left(p^{\mathrm{T}}M^{-1}p - q^{\mathrm{T}}F(q)\right).$$

An extended system method of Andersen [2] includes an external variable V, the volume of the simulation box, as an additional degree of freedom. This mimics the action of a piston on the real system. The kinetic and potential energy for the piston are given by

$$K_{\mathrm{V}} = \frac{1}{2}Q^{-1}p_{\mathrm{V}}^2, \quad U_{\mathrm{V}} = PV$$

where p_{V} is conjugate momentum for V and P is the specified pressure. Introduce fictitious positions and momenta

$$\bar{q} = V^{-1/3}q, \quad \bar{p} = V^{1/3}p$$

and the kinetic and potential energies of the particles become

$$K = \frac{1}{2}V^{-2/3}\bar{p}^{\mathrm{T}}M^{-1}\bar{p}, \quad U = U(V^{1/3}\bar{q}).$$

The Hamiltonian

$$H_{\mathrm{V}} = K + K_{\mathrm{V}} + U + U_{\mathrm{V}}$$

gives rise to the equations of motion

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{q} = V^{-2/3}M^{-1}\bar{p}, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\bar{p} = V^{1/3}F(V^{1/3}\bar{q}),$$
$$\frac{\mathrm{d}}{\mathrm{d}t}V = Q^{-1}p_{\mathrm{V}}, \qquad \frac{\mathrm{d}}{\mathrm{d}t}p_{\mathrm{V}} = P(t) - P$$

where the pressure function $P(t)$ is calculated using

$$P(t) = \frac{1}{3V}\left(V^{-2/3}\bar{p}^{\mathrm{T}}M^{-1}\bar{p} - V^{1/3}\bar{q}^{\mathrm{T}}F(V^{1/3}\bar{q})\right).$$

The trajectories sample the isobaric–isoenthalpic ensemble. Although the system is non-separable, it has some special structure and can be treated explicitly by the generalized leapfrog method.

8 Stochastic Dynamics

Langevin dynamics is an approximation to MD that greatly reduces the number of atoms by omitting explicit solvent atoms and replacing solvent forces by homogenized deterministic and stochastic forces. Brownian dynamics is a further approximation, in which particle masses are set to zero. It speeds up simulations by making longer time steps possible.

Stochastic ordinary differential equations are briefly introduced in appendix B. A good reference on numerical methods for stochastic ODEs is the book [60]. Other references are [53, 61].

8.1 Langevin Dynamics

Simulations of a biomolecule in water spend the bulk of the time doing calculations with water molecules. Hence, it is popular to use an *implicit solvent*, in which the effect of the solvent is replaced by averaged and random forces. By considering N large spherical particles and a number of fluid particles, a second order system of stochastic ODEs known as the Langevin equation can be derived [22].

The Langevin equation is

$$\mathrm{d}q = v\mathrm{d}t, \quad M\mathrm{d}v = F(q)\mathrm{d}t - k_{\mathrm{B}}TD(q)^{-1}v\mathrm{d}t + \sqrt{2}k_{\mathrm{B}}TD_{1/2}(q)^{-\mathrm{T}}\mathrm{d}W(t) \tag{21}$$

where k_{B} is Boltzmann's constant, T is temperature, D is a diffusion tensor, $W(t)$ is a collection of independent standard Wiener processes, and $D_{1/2}$ is a matrix satisfying $D_{1/2}D_{1/2}^{\mathrm{T}} = D$, e.g., a Choleski factorization [24]. In this equation $k_{\mathrm{B}}TD^{-1}$ is the friction tensor. The diffusion tensor D could be non-diagonal; if diagonal, $D_{ii} = (k_{\mathrm{B}}T)/(6\pi\eta a_i)$ where η is the solvent viscosity and a_i the effective radius of the ith particle. Other possible diffusion tensors are given in [24]. Electrostatic effects of the solvent are represented by a potential of mean force incorporated into the force vector F and van der Waals/steric effects are represented by added friction and noise terms. The potential of mean force is obtained from the numerical solution of the Poisson–Boltzmann equation for continuum electrostatics [52].

There are numerous generalizations of the leapfrog method for the Langevin equation. A popular choice is the BBK integrator [12]. Given here is an integrator based on the idea of splitting that is very similar to a discretization proposed in [123]. For convenience we consider the common special case of the Langevin equation (21) where the friction tensor $k_{\mathrm{B}}TD^{-1} = \gamma M$ for some scalar $\gamma \geq 0$:

$$\mathrm{d}q = v\mathrm{d}t, \quad \mathrm{d}v = M^{-1}F(q)\mathrm{d}t - \gamma v\mathrm{d}t + \sqrt{2\gamma k_{\mathrm{B}}T}M^{-1/2}\mathrm{d}W(t).$$

The fractional steps of the integration consist of half a kick, with just the drift term $M^{-1}F(q)\mathrm{d}t$ active, followed by a "fluctuation" with all terms active except the drift term, followed by another half a kick. The method is

half-kick:

$$v^{n+\epsilon} = v^n + \frac{1}{2}\Delta t M^{-1}F^n,$$

fluctuate:

$$\begin{aligned} v^{n+1-\epsilon} &= \mathrm{e}^{-\gamma\Delta t}v^{n+\epsilon} + \sqrt{2\gamma k_{\mathrm{B}}T}M^{-1/2}R_1^{n+1}, \\ q^{n+1} &= q^n + \Delta t\frac{1-\mathrm{e}^{-\gamma\Delta t}}{\gamma\Delta t}v^{n+\epsilon} + \sqrt{2\gamma k_{\mathrm{B}}T}\frac{1}{\gamma}M^{-1/2}R_2^{n+1}, \end{aligned}$$

half-kick:

$$\begin{aligned} F^{n+1} &= F(q^{n+1}), \\ v^{n+1} &= v^{n+1-\epsilon} + \frac{1}{2}\Delta t M^{-1}F^{n+1}. \end{aligned}$$

where

$$R_1^{n+1} = \int_{t^n}^{t^{n+1}} \mathrm{e}^{-\gamma(t^{n+1}-t)}\mathrm{d}W(t), \quad R_2^{n+1} = \int_{t^n}^{t^{n+1}} (1-\mathrm{e}^{-\gamma(t^{n+1}-t)})\mathrm{d}W(t).$$

The pair R_1^{n+1} and R_2^{n+1} are joint Gaussian random variables of zero mean and known covariance matrix C and can be generated as described in appendix B. The elements of the covariance matrix C are the two variances

$$c_{11} = \int_{t^n}^{t^{n+1}} \left(\mathrm{e}^{-\gamma(t^{n+1}-t)}\right)^2 \mathrm{d}t, \quad c_{22} = \int_{t^n}^{t^{n+1}} \left(1-\mathrm{e}^{-\gamma(t^{n+1}-t)}\right)^2 \mathrm{d}t$$

and the covariance

$$c_{12} = c_{21} = \int_{t^n}^{t^{n+1}} \mathrm{e}^{-\gamma(t^{n+1}-t)} \left(1-\mathrm{e}^{-\gamma(t^{n+1}-t)}\right) \mathrm{d}t.$$

We choose $[R_1^{n+1}, R_2^{n+1}]^{\mathrm{T}} = C_{1/2}[Z_1^{n+1}, Z_2^{n+1}]^{\mathrm{T}}$ where Z_1^{n+1}, Z_2^{n+1} are independent Gaussian random numbers of mean 0 and variance 1 and where $C_{1/2}$ is any matrix satisfying $C_{1/2}C_{1/2}^{\mathrm{T}} = C$.

Because of damping, Langevin dynamics is not so sensitive to instabilities, and multiple time stepping is especially advantageous, for example, using the integration method LN [6].

8.2 Brownian Dynamics

The stiffness in the Langevin equations due to the friction term suggests the possibility of a high friction limit approximation. This is attained by setting the masses M to zero in eq. (21) and the resulting equations are those of Brownian dynamics:

$$\mathrm{d}q = \frac{1}{k_{\mathrm{B}}T} D(q)F(q)\mathrm{d}t + \sqrt{2}D_{1/2}(q)\mathrm{d}W(t).$$

This approximation is derived in [24, sec. III]. The spatial scale of the trajectory grows as $\sqrt{t}$ but approximation error from neglecting the inertial term remains bounded as t increases, so Brownian dynamics gives a reasonable approximation over long distances.

The customary choice of numerical integrator is the Euler(–Maruyama) method [60, p. 305], introduced in this context by Ermak and McCammon [24]. This method advances coordinates for a time step Δt by the simple recipe[2]

$$q^{n+1} - q^n = \frac{1}{k_{\mathrm{B}}T} D^n F^n \Delta t + \sqrt{2}D_{1/2}^n \sqrt{\Delta t} Z^{n+1}$$

where Z^{n+1} is a collection of independent random numbers from a Gaussian distribution with mean 0 and variance 1. The time step Δt is chosen so that the force on the particle changes by only a few percent during any step (10% or less according to [19]). The Euler method has *weak* order of accuracy 1 meaning that expected values computed from the trajectories have an error whose magnitude has an expected value of $O(\Delta t)$. There are other stochastic integrators with greater approximation power. The book [60] gives a variety of integrators of higher weak (and strong) order. For MD, weak order is the appropriate concept [60, sec. 17.2].

We conclude with an example illustrating the statistical context of molecular simulations: A number of reactions in the cell have rates that are limited by diffusion, and their rate constant can be calculated by computing large numbers of trajectories using Brownian dynamics. Assuming a dilute solution, only a pair of reactants need to be considered in the calculation. A practical

[2]This can be obtained as the limit as $M \to \infty$ of the method given for Langevin dynamics.

way to get a rate constant from trajectories is the NAM technique, given in [84]. The simulation domain is restricted to intermolecular distances r less than some value r_{cut}, which is chosen large enough so that to a good approximation the probability of reaction beyond r_{cut} depends on only r. Simulations are initiated with the centers of the two reactants at a distance $b < r_{\text{cut}}$ apart where b is sufficiently large that for $r \geq b$ the drift term $F(r)$ depends to a good approximation on only the distance r of separation. Initial values for rotational and internal degrees of freedom are chosen at random from their equilibrium distribution. Each trajectory is simulated until either reaction occurs or separation exceeds r_{cut}. Thousands of trajectories are computed to yield a probability of reaction β. The rate constant can then be calculated from β and a couple of one-dimensional integrals. The time it takes for a particle to reach the end of its trajectory can be arbitrarily long, so it is necessary to put a time limit on the duration of a trajectory and record as "unknown" the value of one which exceeds the time limit.

Acknowledgments

I am grateful to undergraduate student Tony Surma for Figs. 2–5. And to graduate students David Hardy, Jesús Izaguirre, Hui Lu, James Phillips, and Gang Zou for contributing content and feedback that improved the manuscript. And to Ben Leimkuhler for also helping to improve the manuscript.

References

[1] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Clarendon Press, Oxford, New York, 1987. Reprinted in paperback in 1989 with corrections.

[2] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384–2393, 1980.

[3] H. C. Andersen. Rattle: A 'velocity' version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.*, 52:24–34, 1983.

[4] V. I. Arnol'd. *Mathematical Methods of Classical Mechanics*. Springer-Verlag, New York, second edition, 1989.

[5] E. Barth, K. Kuczera, B. Leimkuhler, and R. D. Skeel. Algorithms for constrained molecular dynamics. *J. Comput. Chem.*, 16(10):1192–1209, Oct. 1995.

[6] E. Barth, M. Mandziuk, and T. Schlick. A separating framework for increasing the timestep in molecular dynamics. In W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, editors, *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, volume 3, chapter 4, pages 97–121. ESCOM, Leiden, The Netherlands, 1996.

[7] G. Benettin and A. Giorgilli. On the Hamiltonian interpolation of near to the identity symplectic mappings with application to symplectic integration algorithms. *J. Statist. Phys.*, 74:1117–1143, 1994.

[8] J. J. Biesiadecki and R. D. Skeel. Dangers of multiple-time-step methods. *J. Comput. Phys.*, 109(2):318–328, Dec. 1993.

[9] T. Bishop, R. D. Skeel, and K. Schulten. Difficulties with multiple timestepping and the fast multipole algorithm in molecular dynamics. *J. Comput. Chem.*, 18(14):1785–1791, Nov. 15, 1997.

[10] S. D. Bond, B. J. Leimkuhler, and B. B. Laird. The Nosé–Poincaré method for constant temperature molecular dynamics. Manuscript, 1998.

[11] K. Brenan, S. Campbell, and L. Petzold. *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*. North-Holland, 1989.

[12] A. Brünger, C. B. Brooks, and M. Karplus. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.*, 105:495–500, 1982.

[13] O. Buneman. Time-reversible difference procedures. *J. Comput. Phys.*, 1:517–535, 1967.

[14] J. C. Butcher. The effective order of Runge–Kutta methods. In A. Dold, Z. Heidelberg, and B. Eckmann, editors, *Conference on the Numerical Solution of Differential Equations, Lecture Notes in Mathematics*, volume 109, pages 133–139. Springer-Verlag, New York, 1969.

[15] M. Calvo and J. Sanz-Serna. The development of variable-step symplectic integrators, with applications to the two-body problem. *SIAM J. Sci. Statist. Comput.*, 14:936–952, 1993.

[16] J. Candy and W. Rozmus. A symplectic integration algorithm for separable Hamiltonian functions. *J. Comput. Phys.*, 92:230–256, 1991.

[17] P. J. Channell and J. C. Scovel. Symplectic integration of Hamiltonian systems. *Nonlinearity*, 3:231–259, 1990.

[18] M. M. Chawla. On the order and attainable intervals of periodicity of explicit Nyström methods for $y'' = f(t, y)$. *SIAM J. Numer. Anal.*, 22:127–131, Feb. 1985.

[19] M. E. Davis, J. D. Madura, B. A. Luty, and J. A. McCammon. Electrostatics and diffusion of molecules in solution: Simulations with the University of Houston Brownian dynamics program. *Computer Phys. Comms.*, 62:187–197, 1991.

[20] R. De Vogelaére. Methods of integration which preserve the contact transformation property of Hamiltonian equations. Technical Report 4, Department of Mathematics, University of Notre Dame, 1956.

[21] J. Delambre. *Mem. Acad. Turin*, 5:143, 1790–1793.

[22] J. M. Deutsch and I. Oppenheim. *J. Chem. Phys.*, 54:3547, 1971.

[23] A. Dullweber, B. Leimkuhler, and R. McLachlan. A symplectic splitting method for rigid-body molecular dynamics. *J. Chem. Phys.*, 107:5840, 1997.

[24] D. L. Ermak and J. A. McCammon. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.*, 69(4):1352–1360, Aug. 15, 1978.

[25] D. J. Evans. On the representation of orientation space. *Mol. Phys.*, 34(2):317–325, 1977.

[26] D. J. Evans and G. P. Moriss. Non-Newtonian molecular dynamics. *Comput. Phys. Rep.*, 1:297–344, 1984.

[27] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.*, 103(11):4613–4621, 1995.

[28] K. Feng. On difference schemes and symplectic geometry. In K. Feng, editor, *Proc. 1984 Beijing Symposium on Differential Geometry and Differential Equations—Computation of Differential Equations*, pages 42–58, Science Press, Beijing, 1985.

[29] A. Fischer, F. Cordes, and C. Schütte. Hybrid Monte Carlo with adaptive temperature in a mixed-canonical ensemble: Efficient conformational analysis of rna. Technical Report SC 97-67, Konrad-Zuse-Zentrum für Informationstechnik Berlin, Dec. 1997. Available via http://www.zib.de/bib/pub/pw/.

[30] M. Fixman. Simulation of polymer dynamics. I. General theory. *J. Chem. Phys.*, 69:1527–1537, 1978.

[31] E. Forest and R. D. Ruth. Fourth-order symplectic integration. *Physica D*, 43:105–117, 1990.

[32] T. Forester and W. Smith. On multiple time-step algorithms and the Ewald sum. *Mol. Sim.*, 13(3):195–204, 1994.

[33] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications.* Academic Press, 1996.

[34] B. García-Archilla, J. M. Sanz-Serna, and R. D. Skeel. Long-time-step methods for oscillatory differential equations. *SIAM J. Sci. Comput.* To appear. [Also Tech. Rept. 1996/7, Dep. Math. Aplic. Comput., Univ. Valladolid, Valladolid, Spain].

[35] Z. Ge and J. E. Marsden. Lie–Poisson Hamilton–Jacobi theory and Lie–Poisson integrators. *Phys. Lett. A*, 133(3):134–139, 1988.

[36] C. Gear and D. Wells. Multirate linear multistep methods. *BIT*, 24:484–502, 1984.

[37] C. W. Gear. *Numerical Initial Value Problems in Ordinary Differential Equations.* Prentice-Hall, Englewood Cliffs, N. J., 1971.

[38] D. Goldman and T. J. Kaper. Nth-order operator splitting schemes and nonreversible systems. *SIAM J. Numer. Anal.*, 33:349–367, 1996.

[39] O. Gonzalez and J. Simo. On the stability of symplectic and energy–momentum algorithms for nonlinear Hamiltonian systems with symmetry. *Computer Methods in Applied Mechanics and Engineering*, 134:197–222, 1996.

[40] H. Grubmüller. Dynamiksimulation sehr großer Makromoleküle auf einem Parallelrechner. Master's thesis, Physik-Dept. der Tech. Univ. München, Munich, 1989.

[41] H. Grubmüller. *Molekular dynamik von Proteinen auf langen Zeitskalen.* PhD thesis, Physik-Dept. der Tech. Univ. München, Munich, 1994.

[42] H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation*, 6:121–142, 1991.

[43] J. M. Haile. *Molecular Dynamics Simulation.* John Wiley and Sons, 1992.

[44] E. Hairer. Backward error analysis of numerical integrators and symplectic methods. *Annals of Numer. Math.*, 1:107–132, 1994.

[45] E. Hairer. Variable time step integration with symplectic methods. *Appl. Numer. Math.*, 25(2–3):219–227, Nov. 1997.

[46] E. Hairer and P. Leone. Order barriers for symplectic multi-value methods. In D. Griffiths, D. Higham, and G. Watson, editors, *Proceedings of the 17th Dundee Biennial Conference, June 24–27, 1997*, volume 380 of *Pitman Research Notes in Mathematics*, pages 133–149, 1998.

[47] E. Hairer and C. Lubich. The lifespan of backward error analysis for numerical integrators. *Numer. Math.*, 76:441–462, 1997.

[48] E. Hairer and C. Lubich. Asymptotic expansions and backward analysis for numerical integrators. manuscript, 1998.

[49] D. J. Hardy, D. I. Okunbor, and R. D. Skeel. Symplectic variable stepsize integration for N-body problems. *Appl. Numer. Math.*, 1998. To appear.

[50] A. Hayli. Le problème des N corps dans un champ extérieur application a l'évolution dynamique des amas ouverts-I. *Bulletin Astronomique*, 2:67–89, 1967.

[51] H. Heller, M. Schaefer, and K. Schulten. Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid crystal-phases. *J. Phys. Chem.*, 97:8343–8360, 1993.

[52] M. Holst, R. Kozack, F. Saied, and S. Subramaniam. Treatment of electrostatic effects in proteins: Multigrid-based Newton iterative method for solution of the full nonlinear Poisson–Boltzmann equation. *Proteins: Structure, Function, and Genetics*, 18(3):231–245, 1994.

[53] J. Honerkamp. *Stochastic Dynamical Systems: Concepts, Numerical Methods, Data Analysis*. VCH, 1994.

[54] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.

[55] W. Huang and B. Leimkuhler. The adaptive Verlet method. *SIAM J. Sci. Comput.*, 18:239–256, 1997.

[56] D. D. Humphreys, R. A. Friesner, and B. J. Berne. A multiple-time-step molecular dynamics algorithm for macromolecules. *J. Phys. Chem.*, 98(27):6885–6892, July 7, 1994.

[57] H. Ishida, Y. Nagai, and A. Kidera. Symplectic integrator for molecular dynamics of a protein in water. *Chem. Phys. Letts.*, 282(2):115–120, Jan. 9, 1998.

[58] J. Izaguirre, S. Reich, and R. D. Skeel. Longer time steps for molecular dynamics. Submitted for publication.

[59] D. Janežič and F. Merzel. An efficient symplectic integration algorithm for molecular dynamics simulations. *J. Chem. Inf. Comput. Sci.*, 35:321–326, 1995.

[60] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics: Stochastic Modelling and Applied Probability*. Springer-Verlag, New York, 1992. Second corrected printing 1995.

[61] P. E. Kloeden, E. Platen, and H. Schurz. *Numerical Solution of SDE Through Computer Experiments*. Springer-Verlag, 1994.

[62] A. Kol, B. Laird, and B. Leimkuhler. A symplectic method for rigid-body molecular simulation. *J. Chem. Phys.*, 107(7), 1997.

[63] J. D. Lambert and I. A. Watson. Symmetric multistep methods for periodic initial value problems. *J. Inst. Math. Applics.*, 18:189–202, 1976.

[64] A. Lasota and M. C. Mackey. *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. Springer-Verlag, New York, second edition, 1994.

[65] A. R. Leach. *Molecular Modelling: Principles and Applications*. Addison-Wesley Longman, Reading, Mass., July 1996.

[66] B. Leimkuhler and S. Reich. The numerical solution of constrained Hamiltonian systems. *Math. Comput.*, 63:589–605, 1994.

[67] B. Leimkuhler and R. D. Skeel. Symplectic numerical integrators in constrained Hamiltonian systems. *J. Comput. Phys.*, 112(1):117–125, May 1994.

[68] B. J. Leimkuhler, S. Reich, and R. D. Skeel. Integration methods for molecular dynamics. In J. P. Mesirov, K. Schulten, and D. W. Sumners, editors, *Mathematical Approaches to Biomolecular Structure and Dynamics*, volume 82 of *IMA Volumes in Mathematics and its Applications*, pages 161–185. Springer-Verlag, 1996.

[69] T. R. Littell, R. D. Skeel, and M. Zhang. Error analysis of symplectic multiple time stepping. *SIAM J. Numer. Anal.*, 34(5):1792–1807, Oct. 1997.

[70] M. López-Marcos, J. M. Sanz-Serna, and R. D. Skeel. Cheap enhancement of symplectic integrators. In D. F. Griffiths and G. A. Watson, editors, *Numerical Analysis 1995*, pages 107–122, London, 1996. Longman Group.

[71] M. López-Marcos, J. M. Sanz-Serna, and R. D. Skeel. Explicit symplectic integrators using Hessian–vector products. *SIAM J. Sci. Comput.*, 18:223–238, Jan. 1997.

[72] M. Mandziuk and T. Schlick. Resonance in the dynamics of chemical systems simulated by the implicit midpoint scheme. *Chem. Phys. Letters*, 237:525–535, 1995.

[73] G. J. Martyna. Remarks on 'constant-temperature molecular dynamics with momentum conservation'. *Phys. Rev. E*, 50(4):3234–3236, 1994.

[74] G. J. Martyna, M. L. Klein, and M. Tuckerman. Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.*, 97(5):2635–2643, 1992.

[75] G. J. Martyna, D. J. Tobias, and M. L. Klein. Constant pressure molecular dynamics algorithms. *J. Chem. Phys*, 101(5):4177–4189, Sept. 1, 1994.

[76] R. McLachlan and J. Scovel. Equivariant constrained symplectic integration. *Nonlinear Sci.*, 5:233–256, 1995.

[77] R. I. McLachlan. Explicit Lie–Poisson integration and the Euler equations,. *Phys. Rev. Lett.*, 71:3043–3046, 1993.

[78] R. I. McLachlan. More on symplectic correctors. In J. E. Marsden, G. W. Patrick, and W. F. Shadwick, editors, *Integration Algorithms and Classical Mechanics*, volume 10 of *Fields Institute Communications*. Fields Institute, American Mathematical Society, July 1996. To appear.

[79] R. I. McLachlan and P. Atela. The accuracy of symplectic integrators. *Nonlinearity*, 5:541–562, March 1992.

[80] R. I. McLachlan and S. K. Gray. Optimal stability polynomials for splitting methods, with application to the time-dependent Schrödinger equation. *Appl. Numer. Math.*, 25(2–3):275–286, Nov. 1997.

[81] B. Mehlig, D. W. Heermann, and B. M. Forrest. Hybrid Monte Carlo method for condensed-matter systems. *Phys. Rev. B*, 45(2):679–685, Jan. 1, 1992.

[82] S. Melchionna, G. Ciccotti, and B. L. Holian. Hoover NPT dynamics for systems varying in shape and size. *Mol. Phys.*, 78(3):533–544, 1993.

[83] M. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. Kalé, R. D. Skeel, and K. Schulten. NAMD—a parallel, object-oriented molecular dynamics program. *Intl. J. Supercomput. Applics. High Performance Computing*, 10(4):251–268, Winter 1996.

[84] S. H. Northrup, S. A. Allison, and J. A. McCammon. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.*, 80(4):1517–1524, 1984.

[85] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81:511, 1984.

[86] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer-Verlag, fifth edition, 1998.

[87] D. Okunbor and R. D. Skeel. Explicit canonical methods for Hamiltonian systems. *Math. Comput.*, 59(200):439–455, Oct. 1992.

[88] D. Okunbor and R. D. Skeel. Canonical numerical methods for molecular dynamics simulations. *J. Comput. Chem.*, 15(1):72–79, Jan. 1994.

[89] A. Portillo and J. M. Sanz-Serna. Lack of dissipativity is not symplecticness. *BIT Numer. Math.*, 35(2):269–276, 1995.

[90] G. D. Quinlan and S. Tremaine. Symmetric multistep methods for the numerical integration of planetary orbits. *Astron. J.*, 100:1694–1700, 1990.

[91] S. Reich. Backward error analysis for numerical integrators. To appear.

[92] S. Reich. Momentum preserving symplectic integrators. *Physica D*, 76(4):375–383, Sept. 10, 1994.

[93] S. Reich. A free energy approach to the torsion dynamics of macromolecules. Technical Report SC 95-17, Konrad-Zuse-Zentrum für Informationstechnik Berlin, June 1995.

[94] S. Reich. Smoothed dynamics of highly oscillatory Hamiltonian systems. *Physica D*, 89(1 and 2):28–42, Dec. 21, 1995.

[95] G. Rowlands. A numerical algorithm for Hamiltonian systems. *J. Comput. Phys.*, 97:235–239, Nov. 1991.

[96] R. D. Ruth. A canonical integration technique. *IEEE Trans. Nucl. Sci.*, 30(4):2669–2671, 1983.

[97] J. Sanz-Serna and M. Calvo. *Numerical Hamiltonian Problems*. Chapman and Hall, London, 1994.

[98] J. M. Sanz-Serna. Runge-Kutta schemes for Hamiltonian systems. *BIT*, 28:877–883, 1988.

[99] J. M. Sanz-Serna. Two topics in nonlinear stability. In W. Light, editor, *Advances in Numerical Analysis*, pages 147–174. Clarendon Press, Oxford, 1991.

[100] J. M. Sanz-Serna. Symplectic integrators for Hamiltonian problem: an overview. *Acta Numerica*, 1:243–286, 1992.

[101] T. Schlick, M. Mandziuk, R. D. Skeel, and K. Srinivas. Nonlinear resonance artifacts in molecular dynamics simulations. *J. Comput. Phys.*, 139:1–29, 1998.

[102] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard. A hybrid Monte Carlo method for essential molecular dynamics. Technical Report SC 98-04, Konrad-Zuse-Zentrum für Informationstechnik Berlin, Feb. 1998. Available via http://www.zib.de/bib/pub/pw/.

[103] C. Scovel. On symplectic lattice maps. *Physics Letters A*, 159:396–400, Aug. 1991.

[104] C. Scovel. Symplectic numerical integration of Hamiltonian systems. In T. Ratiu, editor, *The Geometry of Hamiltonian Systems: Proceedings of Workshop Held June 5–16, 1989*, pages 463–496. Mathematical Sciences Research Institute, Springer-Verlag, 1991.

[105] Q. Sheng. *Solving Partial Differential Equations by Exponential Splitting*. PhD thesis, King's College, Cambridge, Oct. 1988.

[106] J. C. Simo and N. Tarnow. A review of conserving algorithms for nonlinear dynamics. Technical Report SUDAM Report 92-7, Dept. Mechanical Engineering, Stanford Univ., Calif., 1992.

[107] R. D. Skeel. Variable step size destabilizes the Störmer/leapfrog/Verlet method. *BIT*, 33:172–175, 1993.

[108] R. D. Skeel. Symplectic integration with floating-point arithmetic and other approximations. *Appl. Numer. Math.*, 1998. To appear.

[109] R. D. Skeel and J. J. Biesiadecki. Symplectic integration with variable stepsize. *Annals of Numer. Math.*, 1:191–198, 1994.

[110] R. D. Skeel and C. W. Gear. Does variable step size ruin a symplectic integrator? *Physica D*, 60:311–313, 1992.

[111] R. D. Skeel and J. Izaguirre. The five femtosecond time step barrier. In P. Deuflhard, J. Hermans, B. Leimkuhler, A. Mark, S. Reich, and R. D. Skeel, editors, *Algorithms for Macromolecular Modelling*, volume 4 of *Lecture Notes in Computational Science and Engineering*, pages 303–318. Springer-Verlag, 1998. To appear.

[112] R. D. Skeel, G. Zhang, and T. Schlick. A family of symplectic integrators: stability, accuracy, and molecular dynamics applications. *SIAM J. Sci. Comput.*, 18(1):203–222, Jan. 1997.

[113] D. M. Stoffer. *Some Geometric and Numerical Methods for Perturbed Integrable Systems*. PhD thesis, Swiss Federal Institute of Technology, Zürich, 1988.

[114] C. Störmer. Sur les trajectoires des corpuscles életrisés. *Arch. Sci.*, 24:5–18, 113–158, 221–247, 1907.

[115] W. B. Streett, D. Tildesley, and G. Saville. Multiple time step methods in molecular dynamics. *Mol. Phys.*, 35:639–648, 1978.

[116] S. J. Stuart, R. H. Zhou, and B. J. Berne. Molecular dynamics with multiple time scales—the selection of efficient reference system propagators. *J. Chem. Phys.*, 105(4):1426—1436, July 22, 1996.

[117] M. Suzuki. General theory of fractal path integrals with applications to many-body theories and statistical physics. *J. Math. Phys.*, 32(2), Feb. 1991.

[118] M. Suzuki. Improved Trotter-like formula. *Physics Letters A*, 180(3), June 1993.

[119] D. J. Tobias, G. J. Martyna, and M. L. Klein. Molecular dynamics simulations of a protein in the canonical ensemble. *J. Phys. Chem.*, 97(47):12959–12966, 1993.

[120] S. Toxvaerd. Hamiltonians for discrete dynamics. *Phys. Rev. E*, 1995.

[121] M. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *J. Chem. Phys*, 97(3):1990–2001, 1992.

[122] W. F. van Gunsteren and H. J. C. Berendsen. Algorithms for macromolecular dynamics and constraint dynamics. *Molecular Phys*, 34:1311–1327, 1977.

[123] W. F. van Gunsteren and H. J. C. Berendsen. A leap-frog algorithm for stochastic dynamics. *Molecular Simulation*, 1987. To appear.

[124] W. F. van Gunsteren and H. J. C. Berendsen. *GROMOS Manual*. BIOMOS b. v., Lab. of Phys. Chem., Univ. of Groningen, 1987.

[125] L. Verlet. Computer 'experiments' on condensed fluids I. thermodynamical properties of Lennard–Jones molecules. *Phys. Rev.*, 159:98–103, 1967.

[126] R. F. Warming and B. J. Hyett. The modified equation approach to the stability and accuracy analysis of finite difference methods. *J. Comput. Phys*, 14:159–179, 1974.

[127] M. Watanabe and M. Karplus. Simulation of macromolecules by multiple-time-step methods. *J. Phys. Chem.*, 99(15):5680–5697, Apr. 13, 1995.

[128] J. Wisdom. The origin of the Kirkwood gaps: A mapping for asteroidal motion near the 3/1 commensurability. *Astr. J.*, 87:577–593, 1982.

[129] J. Wisdom, M. Holman, and J. Touma. Symplectic correctors. In J. E. Marsden, G. W. Patrick, and W. F. Shadwick, editors, *Integration Algorithms and Classical Mechanics*, volume 10 of *Fields Institute Communications*. Fields Institute, American Mathematical Society, July 1996. To appear.

[130] H. Yoshida. Construction of higher order symplectic integrators. *Phys. Lett. A*, 150:262–268, 1990.

[131] M. Q. Zhang and R. D. Skeel. Symplectic integrators and the conservation of angular momentum. *J. Comput. Chem.*, 16:365–369, Mar. 1995.

A Lie Series and the BCH Formula

The main purpose of this appendix is to obtain a formula for the composition of Hamiltonian flows based on the Baker–Campbell–Hausdorff (BCH) formula for Lie operators. The details are given below. The conclusion is the following formula:

$$\phi_{(\Delta t/2)N} \circ \phi_{\Delta t H} \circ \phi_{(\Delta t/2)N} = \phi_{\Delta t H^+}$$

where

$$\begin{aligned} H^+ &= H + N + \frac{\Delta t^2}{12}\{H,H,N\} - \frac{\Delta t^2}{24}\{N,N,H\} + \frac{\Delta t^4}{5760}\{N,N,N,N,H\} - \frac{\Delta t^4}{720}\{H,H,H,H,N\} \\ &\quad + \frac{\Delta t^4}{360}\{N,H,H,H,N\} + \frac{\Delta t^4}{360}\{H,N,N,N,N\} - \frac{\Delta t^4}{480}\{N,N,H,H,N\} + \frac{\Delta t^4}{120}\{H,H,N,N,H\} \\ &\quad + O(\Delta t^6). \end{aligned}$$

The Poisson bracket of two scalar fields is defined by

$$\{H,N\} = H_y^{\mathrm{T}} J N_y.$$

The first step is to express the flow defined by a general ODE system as the exponential of a Lie operator. Let $y(t)$ be a trajectory for an ODE system $(\mathrm{d}/\mathrm{d}t)y = f(y)$. Let $g(y)$ be a scalar function of y. Normally for small enough Δt one can write

$$g(y(t+\Delta t)) = \sum_{k=0}^{\infty} \frac{\Delta t^k}{k!} \frac{\mathrm{d}^k}{\mathrm{d}t^k} g(y(t)). \tag{22}$$

The idea is to use the ODE to generate the first and higher derivatives. Differentiating along a trajectory gives

$$\frac{\mathrm{d}}{\mathrm{d}t} g(y(t)) = \frac{\mathrm{d}}{\mathrm{d}t} y(t)^{\mathrm{T}} g_y(y(t)) = ((f^{\mathrm{T}}\nabla)g)(y(t)) \tag{23}$$

where the Lie operator $f^{\mathrm{T}}\nabla$ is defined by $(f^{\mathrm{T}}\nabla)g(y) = f(y)^{\mathrm{T}} g_y(y)$. It is straightforward to show by induction that

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k} g(y(t)) = ((f^{\mathrm{T}}\nabla)^k g)(y(t))$$

where $(f^{\mathrm{T}}\nabla)^k$ is defined recursively by $(f^{\mathrm{T}}\nabla)^k g = (f^{\mathrm{T}}\nabla)((f^{\mathrm{T}}\nabla)^{k-1} g)$. Substituting the above into (22), we get

$$g(y(t+\Delta t)) = (\exp(\Delta t f^{\mathrm{T}}\nabla)g)(y(t)). \tag{24}$$

We generalize from $g(y)$ to a vector field by defining $f^{\mathrm{T}}\nabla$ component-wise. For the vector field $\mathrm{id}(y)$, eq. (24) gives the formula for the flow:

$$y(t+\Delta t) = (\exp(\Delta t f^{\mathrm{T}}\nabla)\mathrm{id})(y(t)).$$

Therefore, eq. (24) can be written

$$g \circ (\exp(\Delta t f^{\mathrm{T}}\nabla)\mathrm{id}) = \exp(\Delta t f^{\mathrm{T}}\nabla) g. \tag{25}$$

Next, we explain how the BCH formula can be used to compose flows defined by general ODE systems. Two applications of the identity (25) yield

$$\begin{aligned} g \circ \exp(\Delta t f_2^{\mathrm{T}}\nabla)\mathrm{id} \circ \exp(\Delta t f_1^{\mathrm{T}}\nabla)\mathrm{id} &= \exp(\Delta t f_1^{\mathrm{T}}\nabla)(g \circ \exp(\Delta t f_2^{\mathrm{T}}\nabla)\mathrm{id}) \\ &= \exp(\Delta t f_1^{\mathrm{T}}\nabla)\exp(\Delta t f_2^{\mathrm{T}}\nabla) g, \end{aligned}$$

which gives us a way to write a composition of flows as a product of exponentials:

$$\exp(\Delta t f_k^{\mathrm{T}}\nabla)\mathrm{id}\circ\cdots\circ\exp(\Delta t f_2^{\mathrm{T}}\nabla)\mathrm{id}\circ\exp(\Delta t f_1^{\mathrm{T}}\nabla)\mathrm{id} = \exp(\Delta t f_1^{\mathrm{T}}\nabla)\exp(\Delta t f_2^{\mathrm{T}}\nabla)\cdots\exp(\Delta t f_k^{\mathrm{T}}\nabla)\mathrm{id}.$$

A product of exponentials can be combined using the Baker–Campbell–Hausdorff formula. Of most use to us is the special case of a symmetric product [97, p. 161]:

$$\exp(\frac{1}{2}B)\exp(A)\exp(\frac{1}{2}B) = \exp(C)$$

where

$$\begin{aligned} C &= A + B + \frac{1}{12}[A,A,B] - \frac{1}{24}[B,B,A] + \frac{7}{5760}[B,B,B,B,A] - \frac{1}{720}[A,A,A,A,B] \\ &\quad + \frac{1}{360}[B,A,A,A,B] + \frac{1}{360}[A,B,B,B,A] - \frac{1}{480}[B,B,A,A,B] + \frac{1}{120}[A,A,B,B,A] + \cdots. \end{aligned}$$

The commutator of two operators is defined by $[A,B] = AB - BA$, and the iterated commutator is defined recursively by $[A_1, A_2, \ldots, A_k] = [A_1, [A_2, \ldots, A_k]]$.

Finally, we specialize the BCH formula to the case of a Hamiltonian flow $f(y) = JH_y(y)$ by noting the following homomorphism from scalar fields with the negative Poisson bracket to Lie operators with the commutator:

$$[(JH_y)^{\mathrm{T}}\nabla, (JN_y)^{\mathrm{T}}\nabla] = (J(-\{H,N\})_y)^{\mathrm{T}}\nabla.$$

B Stochastic Processes

Given here is some information about stochastic processes helpful for understanding the derivation of simple numerical integrators. Books on stochastic processes include [86, 60].

B.1 Wiener processes

A *continuous time stochastic process* is a family of random variables $y(t)$ indexed for some continuous range of parameter values $0 \le t \le T$. For any outcome, $y(t)$ will be a function of t called a *realization* or *sample path* of the process. Many useful stochastic processes can be constructed from *Wiener processes*, and we restrict our attention to such processes.

A standard Wiener process $W(t)$, $t \ge 0$, can be defined [60, p. 28] as follows:

1. $W(0) = 0$ with probability 1 (w. p. 1),
2. $W(t+\Delta t) - W(t)$, $\Delta t > 0$, is independent of $W(\tau)$ for $\tau \le t$ and is Gaussian with mean zero and variance Δt.

Hence, the increment $Y = W(t+\Delta t) - W(t)$ has probability density function

$$\frac{1}{\sqrt{2\pi\Delta t}}\exp\left(-\frac{y^2}{2\Delta t}\right).$$

To make this more concrete, three methods for generating a Wiener process are given below. In all cases Z^n designates a sequence of independent standard Gaussian random variables (with mean 0 and variance 1).

1. $W(0) = 0$, $W(t^n) = W(t^{n-1}) + \sqrt{t^n - t^{n-1}}Z^n$.

2. In the case where $W(t)$ is being evaluated at a sequence of arbitrary points not known in advance and not necessarily increasing (such as might occur if it were being plotted by an adaptive procedure), then use

$$W(t^n) = \frac{b^n - t^n}{b^n - a^n} W(a^n) + \frac{t^n - a^n}{b^n - a^n} W(b^n) + \sqrt{\frac{(t^n - a^n)(b^n - t^n)}{b^n - a^n}} Z^n$$

where a^n is the point from the set $\{0, t^1, \ldots, t^{n-1}\}$ immediately to the left of t^n and b^n is the point from the same set immediately to the right of t^n.

3. The Karhunen-Loève expansion, good for a finite interval $[0, T]$, is

$$\sum_{n=0}^{\infty} Z^n \frac{2\sqrt{2T}}{(2n+1)\pi} \sin\left(\frac{(2n+1)\pi t}{2T}\right).$$

These constructions indicate that a realization of $W(t)$ is continuous but *not differentiable* at any given point t w. p. 1.

B.2 The Ito integral

The construction of useful stochastic processes from the Wiener process requires integrals of the form

$$I(f) = \int_0^T f(t) \mathrm{d}W(t)$$

For a differentiable deterministic function $f(t)$ we can use the Riemann–Stieltjes integral, which is the limit as $\Delta t \to 0$ of

$$\sum_{n=1}^{N} f(\tau^n)(W(t^n) - W(t^{n-1}))$$

where $0 = t^0 < t^1 < \cdots < t^N = T$, $\Delta t = \max_n(t^n - t^{n-1})$, and each τ^n is an arbitrary value in the closed interval from t^{n-1} to t^n. We have that

- $I(f)$ is a Gaussian random variable,
- $E[I(f)] = 0$,
- $E[I(f)^2] = \int_0^T f(t)^2 \mathrm{d}t$,
- $E[I(f)I(g)] = \int_0^T f(t)g(t)\mathrm{d}t$.

We can generate $\int_0^T f(t)\mathrm{d}W(t)$ as $(\int_0^T f(t)^2\mathrm{d}t)^{1/2}Z$. If other integrals $\int_0^T g(t)\mathrm{d}W(t)$ are also to be generated, the covariance of these random variables has to be taken into account. Jointly Gaussian random variables Y_1, Y_2, ..., Y_k with mean vector μ and covariance matrix C have joint p.d.f.

$$\rho(y) = (2\pi \det C)^{-k/2} \exp(-\frac{1}{2}(y-\mu)^{\mathrm{T}} C (y-\mu)).$$

The following formula generates appropriately random values:

$$Y = C_{1/2} Z + \mu$$

where Z is a vector of independent standard Gaussian random variables and $C_{1/2}$ is any matrix satisfying $C_{1/2} C_{1/2}^{\mathrm{T}} = C$, for example, a Choleski factor.

The Riemann–Stieltjes integral is inadequate for a non-differentiable stochastic integrand such as $W(t)$, because the result of the limit depends on the choice of the τ^n. The choice $\tau^n = t^{n-1}$, where the integrand is evaluated at the very beginning of each subinterval, is known as the Ito integral. It is computationally convenient [86, p. 37], [60, p. 228] because it is a *martingale* and yields an integral with expected value = 0, although it may not be the most physically reasonable choice [60, p. 228]. For the integrand $W(t)$ the Ito integral is

$$\begin{aligned}\int_0^T W(t)\mathrm{d}W(t) &= \lim_{\Delta t\to 0}\sum_{n=1}^{N} W(t^{n-1})(W(t^n)-W(t^{n-1})) \\ &= \lim_{\Delta t\to 0}\sum_{n=1}^{N}\left(\frac{1}{2}W(t^n)^2-\frac{1}{2}W(t^{n-1})^2-\frac{1}{2}(W(t^n)-W(t^{n-1}))^2\right) \\ &= \frac{1}{2}W(T)^2-\lim_{\Delta t\to 0}\frac{1}{2}\sum_{n=1}^{N}(t^n-t^{n-1})(Z^n)^2,\end{aligned}$$

where the Z^n are independent Gaussian random variables, for which $\mathrm{E}((Z^n)^2)=1$ and $\mathrm{Var}((Z^n)^2)=2$. Hence, we have

$$\begin{aligned}\mathrm{E}\left(\frac{1}{2}\sum_{n=1}^{N}(t^n-t^{n-1})(Z^n)^2\right) &= \frac{1}{2}T, \\ \mathrm{Var}\left(\frac{1}{2}\sum_{n=1}^{N}(t^n-t^{n-1})(Z^n)^2\right) &= \sum_{n=1}^{N}(t^n-t^{n-1})^2 \le \Delta t T.\end{aligned}$$

Therefore,

$$\int_0^T W(t)\mathrm{d}W(t) = \frac{1}{2}W(T)^2-\frac{1}{2}T \quad \text{w.p.1.} \tag{26}$$

B.3 Stochastic differential equations

It is by means of stochastic differential equations that most interesting stochastic processes are generated. The general stochastic ODE has the form

$$\mathrm{d}y(t) = f(y(t),t)\mathrm{d}t + B(y(t),t)\mathrm{d}W(t), \tag{27}$$

a drift term plus a diffusion term. Because $W(t)$ is not differentiable, this should be understood as an integral equation. Perhaps the easiest way to understand its meaning is as the limit as $\Delta t\to 0$ of the Euler–Maruyama discretization given in section 8.2.

B.4 The Fokker–Planck equation

Let $\rho(y,t)$ be the p.d.f. of $y(t)$:

$$\Pr(y(t)\in\Omega) = \int_\Omega \rho(y,t)\mathrm{d}y.$$

It satisfies the following generalization of the Liouville equation:

$$\rho_t + \nabla^{\mathrm{T}}(\rho f - \frac{1}{2}(\nabla^{\mathrm{T}}(\rho B B^{\mathrm{T}}))^{\mathrm{T}}) = 0$$

where $f = f(y,t)$ and $B = B(y,t)$.

B.5 The Ito formula

The *Ito formula* [60, pp. 92, 97], [86, pp. 44, 49] provides for the Ito integral the needed generalizations of the fundamental theorem of calculus and of (an integral formulation of) the chain rule for differentiation. Following is a Lie operator version of the Ito formula [60, pp. 163, 177], which generalizes eq. (23) to a stochastic process $y(t)$ defined by the stochastic ODE (27):

$$g(y(T),T) = g(y(0),0) + \int_0^T (L^0 g)(y(t),t)\mathrm{d}t + \int_0^T ((Lg)(y(t),t))^{\mathrm{T}}\mathrm{d}W(t) \tag{28}$$

where $g(y,t)$ is any smooth scalar function,

$$\begin{aligned} L^0 g(y,t) &= g_t(y,t) + g_y(y,t)^{\mathrm{T}} f(y,t) + \frac{1}{2}\mathrm{tr}\left(B(y,t)^{\mathrm{T}} g_{yy}(y,t) B(y,t)\right), \\ (Lg(y,t))^{\mathrm{T}} &= g_y(y,t)^{\mathrm{T}} B(y,t), \end{aligned}$$

and tr denotes the trace. A more heuristic formulation of the Ito formula, easier to remember, is given by

$$\mathrm{d}g = g_t\mathrm{d}t + g_y^{\mathrm{T}}\mathrm{d}y + \frac{1}{2}(\mathrm{d}y)^{\mathrm{T}} g_{yy}(\mathrm{d}y)$$

where

$$\mathrm{d}y = f\mathrm{d}t + B\mathrm{d}W(t)$$

and

$$\mathrm{d}t\mathrm{d}t = 0, \quad \mathrm{d}W(t)\mathrm{d}t = 0, \quad \mathrm{d}W(t)(\mathrm{d}W(t))^{\mathrm{T}} = \mathrm{d}tI.$$

The Ito formula for the special case of a Wiener process $y(t) = W(t)$ ($f = 0$, $B = I$) is

$$g(W(T),T) = g(0,0) + \int_0^T \left(g_t(W(t),t) + \frac{1}{2}tr(g_{yy}(W(t),t))\right)\mathrm{d}t + \int_0^T g_y(W(t),t)^{\mathrm{T}}\mathrm{d}W(t).$$

As examples, $g(y) = \frac{1}{2}y^2$ gives (26), and $g(y) = ty$ gives

$$TW(T) = \int_0^T W(t)\mathrm{d}t + \int_0^T t\mathrm{d}W(t).$$

B.6 Weak Ito–Taylor expansions

The Lie series expansion of a solution to an ODE given in appendix A generalizes to the stochastic ODE (27) through the repeated use of the Ito formula (28). This avoids the direct differentiation of a stochastic process $y(t)$ with respect to time. For simplicity we consider the case of a scalar stochastic ODE. Application of the Ito formula (28) to $g(y,t) = y$ yields

$$y(T) = y(0) + \int_0^T (L^0\mathrm{id})(y(t),t)\mathrm{d}t + \int_0^T (L\mathrm{id})(y(t),t)^{\mathrm{T}}\mathrm{d}W(t)$$

where

$$L^0\mathrm{id} = f, \quad L\mathrm{id} = B.$$

Substituting the Ito formula for $g(y,t) = (L^0\mathrm{id})(y,t)$ and for $g(y,t) = (L\mathrm{id})(y,t)$ into this equation gives

$$\begin{aligned} y(T) &= y(0) + \int_0^T \left((L^0\mathrm{id})(y(0),0) + \int_0^t (L^0L^0\mathrm{id})(y(\tau),\tau)\mathrm{d}\tau + \int_0^t (LL^0\mathrm{id})(y(\tau),\tau)\mathrm{d}W(\tau)\right)\mathrm{d}t \\ &\quad + \int_0^T \left((L\mathrm{id})(y(0),0) + \int_0^t (L^0L\mathrm{id})(y(\tau),\tau)\mathrm{d}\tau + \int_0^t (LL\mathrm{id})(y(\tau),\tau)\mathrm{d}W(\tau)\right)\mathrm{d}W(t) \end{aligned}$$

where

$$L^0L^0\mathrm{id} = f_t + ff_y + \frac{1}{2}B^2 f_{yy}, \quad LL^0\mathrm{id} = Bf_y,$$

$$L^0L\mathrm{id} = B_t + fB_y + \frac{1}{2}B^2 B_{yy}, \quad LL\mathrm{id} = BB_y.$$

If we carry the expansion one step further and discard the triple integral remainer terms, we get a truncated Ito-Taylor expansion of weak order 2 [60, p. 211]:

$$\begin{aligned} y(T) &= y(0) + (L^0\mathrm{id})(y(0),0)\int_0^T \mathrm{d}t + (L\mathrm{id})(y(0),0)\int_0^T \mathrm{d}W(t) \\ &\quad + (L^0L^0\mathrm{id})(y(0),0)\int_0^T\int_0^t \mathrm{d}\tau\mathrm{d}t + (LL^0\mathrm{id})(y(0),0)\int_0^T\int_0^t \mathrm{d}W(\tau)\mathrm{d}t \\ &\quad + (L^0L\mathrm{id})(y(0),0)\int_0^T\int_0^t \mathrm{d}\tau\mathrm{d}W(t) + (LL\mathrm{id})(y(0),0)\int_0^T\int_0^t \mathrm{d}W(\tau)\mathrm{d}W(t) \end{aligned}$$

Weak order of accuracy 2 means that expected values computed from the trajectories have an error of expected magnitude $O(\Delta t^2)$. To use this a numerical method, note that

$$\begin{aligned} \int_0^T \mathrm{d}t &= T, \\ \int_0^T \mathrm{d}W(t) &=: R_1, \\ \int_0^T\int_0^t \mathrm{d}\tau\mathrm{d}t &= \frac{1}{2}T^2, \\ \int_0^T\int_0^t \mathrm{d}W(\tau)\mathrm{d}t = TW(T) - \int_0^T t\mathrm{d}W(t) &=: TR_1 - R_2, \\ \int_0^T\int_0^t \mathrm{d}\tau\mathrm{d}W(t) = \int_0^T t\mathrm{d}W(t) &= R_2, \\ \int_0^T\int_0^t \mathrm{d}W(\tau)\mathrm{d}W(t) = \frac{1}{2}W(T)^2 - \frac{1}{2}T^2 &= \frac{1}{2}R_1^2 - \frac{1}{2}T^2 \end{aligned}$$

where R_1 and R_2 can be computed as explained in sec. 8.1. It is possible to obtain both R_1 and R_2 with a single random number [60, p. XXXII] (when weak convergence is sufficient).