STRUCTURE PRESERVING INTEGRATION ALGORITHMS*

B. A. SHADWICK[†][‡][§], WALTER F. BUELL[¶][†][||], AND JOHN C. BOWMAN[**]

Abstract. Often in physics and engineering one encounters systems of differential equations that have a non-trivial dynamic or kinematic structure, *e.g.*, the flow generated by such a system may satisfy one or more algebraic or differential constraints. Moreover, this structure is often of physical significance, embodying an important concept such as conservation of energy. Traditional numerical methods for solving initial value problems typically do not preserve any structure possessed by the system and can be computationally less efficient than algorithms specifically designed to honour a system's structure. Also of interest are "near ideal" systems, where some conservation property is only weakly violated. Through a series of examples drawn from various physical systems, we discuss numerical algorithms which, in each case, are specifically constructed to preserve the structure of the system under consideration. These methods are shown to be of particular interest when the integration interval is significantly longer than the characteristic time scale(s) of the system.

Key words. conservative, integration, numerical, symplectic

AMS subject classifications. 65L05, 34-04, 34A50

1. Introduction.

In physics and engineering one often encounters systems described by differential equations which possess a non-trivial *structure* that embodies important physical properties which can affect the *qualitative* behaviour of the system. We use the term "structure" in a broad sense — we consider any constraint upon the solutions as structure. Systems of ODEs can possess a wide range of structure; for example, in classical mechanics the Poincaré differential invariants result in conservation of (projections of) phase space volume. In addition, systems may possess constants of motion (first integrals) such as energy or angular momentum, *etc.* Systems described by PDEs have the extra complication that they admit the possibility of an (uncountable) infinity of invariants such as the Casimir invariants found in continuum mechanics [11]. For example, any functional of vorticity is conserved by Euler's equations, likewise any functional of the phase-space density is conserved by the Vlasov equation.

For systems with structure, all numerical errors are not equal. One can think of the system's structure as restricting the dynamical variables to some n-dimensional surface [7] and then separate numerical errors into two categories: those *tangent* to this surface and those *normal* to this surface. As these errors accumulate over many time steps, those in the latter category are errors that violate the structure of the system (for example, they may represent energy gain or dissipation in a conservative system) while those in the former category are more benign as they represent quantitative rather than qualitative errors.

There is much anecdotal evidence (see for example the discussion in Refs. [8, 13]) that numerical methods which preserve the structure of a system are likely to yield results superior to (more accurate than) a generic method, or, alternatively, that for a

[†] The Institute for Advanced Physics, 10875 U.S. Hwy. 285, Suite 199, Conifer, Colorado, 80433.

[‡] Center for Beam Physics, Lawrence Berkeley National Laboratory, Berkeley, CA 94720-0001.

[§] Email: BAShadwick@IAPhysics.org.

[¶] Electronics and Photonics Laboratory, The Aerospace Corporation, M2-253, P.O. Box 92957, Los Angeles, CA 90009-2957

[||] Email: Walter.F.Buell@aero.org.

[**] Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1 (bowman@math.ualberta.ca).

[*] Supported in part by the Institute for Advanced Physics.

given accuracy, a structure preserving method is likely to require less computational effort. This is in part due to structure-preserving methods having better stability properties than generic methods; in some cases structure preserving methods can exhibit unconditional *non-linear* stability. (Furthermore, for the Korteweg–de Vries equation it has been rigorously shown [8] that conservative methods exhibit less rapid error growth than do non-conserving methods.) This seems to be especially true in cases where the time domain of interest is much larger than the system's characteristic time scale(s).

These concepts are also applicable to weakly non-ideal systems, *i.e.*, those systems that can be viewed as a perturbation to an ideal system. For these systems the evolution is obtained using operator splitting; one separates the differential operator into a piece describing the ideal system and a piece describing the perturbation. One uses a structure-preserving method for the ideal part and a generic method for the perturbation. The complete evolution is obtained by combining the evolution for the separate operators and has the desirable property that in the limit as the perturbation vanishes, the structure-preserving behaviour is recovered [4, 14].

We will proceed by considering several different examples of structure preserving methods drawn from various branches of physics.

2. Symplectic Integrators. Consider a one-dimensional Hamiltonian system:

$$\dot{q} = -\frac{\partial H}{\partial p} \quad \text{and} \quad \dot{p} = \frac{\partial H}{\partial q}\,. \tag{1}$$

(Throughout, an over-dot will be used to denote a time derivative.) Since this is a Hamiltonian system, phase space volume is conserved and thus the Poisson bracket of q and p at any time is unity:

$$[q(t), p(t)]_{\{q(t'),p(t')\}} \equiv \frac{\partial q(t)}{\partial q(t')}\frac{\partial p(t)}{\partial p(t')} - \frac{\partial p(t)}{\partial q(t')}\frac{\partial q(t)}{\partial p(t')} = 1. \tag{2}$$

That is, the dynamical variables at any one time are related to the dynamical variables at another time by a canonical transformation. Now, suppose one uses Euler's method to numerically evolve this system. Denoting the numerical solution at time $t_n = n\,\tau$ by $\mathfrak{q}^n$ and $\mathfrak{p}^n$, where the derivatives of H are evaluated at $\mathfrak{q}^n$ and $\mathfrak{p}^n$,

$$\begin{aligned} \mathfrak{q}^{n+1} &= \mathfrak{q}^n - \tau\,\frac{\partial H}{\partial p}(\mathfrak{q}^n, \mathfrak{p}^n), \\ \mathfrak{p}^{n+1} &= \mathfrak{p}^n + \tau\,\frac{\partial H}{\partial q}(\mathfrak{q}^n, \mathfrak{p}^n). \end{aligned} \tag{3}$$

The pertinent question is "Does this time-advance map represent a canonical transformation?" A straightforward calculation reveals

$$[\mathfrak{q}^{n+1}, \mathfrak{p}^{n+1}]_{\{\mathfrak{q}^n,\mathfrak{p}^n\}} = 1 - \tau^2\left(\left[\frac{\partial^2 H}{\partial q\,\partial p}\right]^2 - \frac{\partial^2 H}{\partial q^2}\frac{\partial^2 H}{\partial p^2}\right), \tag{4}$$

where again the derivatives of H are evaluated at $\mathfrak{q}^n$ and $\mathfrak{p}^n$. We see that evolution given by (3) is not Hamiltonian, *i.e.,* it will not preserve phase space volume.

The solution, as was originally recognized by De Vogelære [9] in the mid 1950's, is to make the numerical approximation to the evolution a canonical transformation.

The net result is that the numerical evolution is the *exact solution* of some Hamiltonian system that approximates the original system. In this way, the numerical error normal to the constraint surface representing the Poincaré invariants is eliminated. The importance of this method for numerically evolving Hamiltonian systems was rediscovered in the 1980's by accelerator physicists when designing the Superconducting Super Collider, where it was necessary to understand the smearing of phase space[1] due to non-linearities in the machine over 10^9 particle orbits. In this case, symplectic methods were necessary both to keep the computational demands reasonable and also to ensure that the phase space distortions seen were solely a consequence of the machine design and not numerical artifacts. Symplectic methods also figured prominently in the pioneering work of Wisdom and coworkers in exploring the long-term stability of the solar system, see for example Ref. [16].

For a comprehensive review of these techniques, see the article by Channel and Scovel [5] as well as the monograph by Sans-Serna and Calvo [12]. A large variety of methods, both explicit (which are based largely on operator splitting [17]) and implicit are now known. As a simple example, the Mid-Point rule (where the equations of motion are differenced *between* time steps),

$$
\begin{aligned}
\mathfrak{q}^{n+1} &= \mathfrak{q}^n - \tau \frac{\partial H}{\partial p}\left(\frac{\mathfrak{q}^{n+1}+\mathfrak{q}^n}{2}, \frac{\mathfrak{p}^{n+1}+\mathfrak{p}^n}{2}\right), \\
\mathfrak{p}^{n+1} &= \mathfrak{p}^n + \tau \frac{\partial H}{\partial q}\left(\frac{\mathfrak{q}^{n+1}+\mathfrak{q}^n}{2}, \frac{\mathfrak{p}^{n+1}+\mathfrak{p}^n}{2}\right).
\end{aligned} \tag{5}
$$

yields a time-advance map that is a canonical transformation.

3. Exactly Conservative Integrators. We now move on to consider spectral truncations of the Euler equations. In particular, we consider the "three-wave" problem obtained by restricting the Fourier-transformed equations to three modes [1, 2, 3, 6, 13]:

$$
\begin{aligned}
\dot{\phi}_K &= M_K\,\phi_P\,\phi_Q \equiv S_K(\phi)\,, \\
\dot{\phi}_P &= M_P\,\phi_Q\,\phi_K \equiv S_P(\phi)\,, \\
\dot{\phi}_Q &= M_Q\,\phi_K\,\phi_P \equiv S_Q(\phi)\,,
\end{aligned} \tag{6}
$$

where $\phi = (\phi_K, \phi_P, \phi_Q)$, K, P, and Q are the magnitudes of the Fourier wavenumbers of the three modes, and the coefficients M_K, M_P, and M_Q satisfy

$$M_K + M_P + M_Q = 0 \tag{7}$$

and

$$K^2\,M_K + P^2\,M_P + Q^2\,M_Q = 0\,. \tag{8}$$

This system possesses two invariants: the total energy

$$E = \frac{1}{2}\left(\phi_K^2 + \phi_P^2 + \phi_Q^2\right) \tag{9}$$

[1] Minimizing the phase space occupied by the beam ("emittance" in the language of accelerator physicists) is critical to obtaining high luminosity and consequently to the usefulness of the collider for physics experiments.

and the total enstrophy

$$Z = \frac{1}{2}\left(K^2\,\phi_K^2 + P^2\,\phi_P^2 + Q^2\,\phi_Q^2\right). \tag{10}$$

These invariants follow directly from properties of S_k:

$$\sum_{k\in\{K,P,Q\}} \phi_k\,S_k = 0\,, \tag{11a}$$

$$\sum_{k\in\{K,P,Q\}} k^2\,\phi_k\,S_k = 0\,. \tag{11b}$$

(The equations (6), are identical to Euler's equations for the rigid body, in which case the additional invariant is the norm of the total angular momentum.)

When (6) is integrated numerically using standard explicit methods, neither E nor Z are exactly conserved. This can be illustrated by applying Euler's method. Denoting the numerical solution at time $t_n = n\,\tau$ by φ^n,

$$\varphi_k^{n+1} = \varphi_k^n + \tau\,S_k(\varphi^n)\,; \qquad k\in\{K,P,Q\}\,. \tag{12}$$

The energy at the new time is

$$\begin{aligned} E(\varphi^{n+1}) &= \frac{1}{2}\sum_k\left[\varphi_k^n + \tau\,S_k(\varphi^n)\right]^2 \\ &= E(\varphi^n) + \frac{1}{2}\tau^2\sum_k S_k(\varphi^n)^2\,, \end{aligned} \tag{13}$$

where we have used (11a) in the last step. Thus the total energy is *always* increasing. Similarly we find

$$Z(\varphi^{n+1}) = Z(\varphi^n) + \frac{1}{2}\tau^2\sum_k k^2\,S_k(\varphi^n)^2\,, \tag{14}$$

which is likewise always increasing.

Inspired by the idea of backwards error analysis, one is left to wonder if it possible to modify the equations of motion such that, for a given integrator, the time-advance mapping conserves energy and enstrophy while still yielding a numerical approximation to the original system [13]. More formally, consider the modified system

$$\dot{\phi}_k = S_k(\phi) + f_k\,, \tag{15}$$

where f_k is chosen to guarantee *exact* conservation of energy and enstrophy and to vanish in the limit $\tau \to 0$ sufficiently rapidly that the numerical solution is still consistent with the original system (*i.e.,* the order of the new time-advance map should be the same as the original integrator).

We illustrate this analysis using a second-order predictor–corrector algorithm (Heun's method [18]). Applying this integrator to the modified system we have

$$\begin{aligned} \widetilde{\varphi}_k &= \varphi_k^n + \tau\left(S_k(\varphi^n) + \mathfrak{f}_k\right), \\ \varphi_k^{n+1} &= \varphi_k^n + \frac{\tau}{2}\left(S_k(\varphi^n) + \mathfrak{f}_k + \widetilde{S}_k + \widetilde{\mathfrak{f}}_k\right), \end{aligned} \tag{16}$$

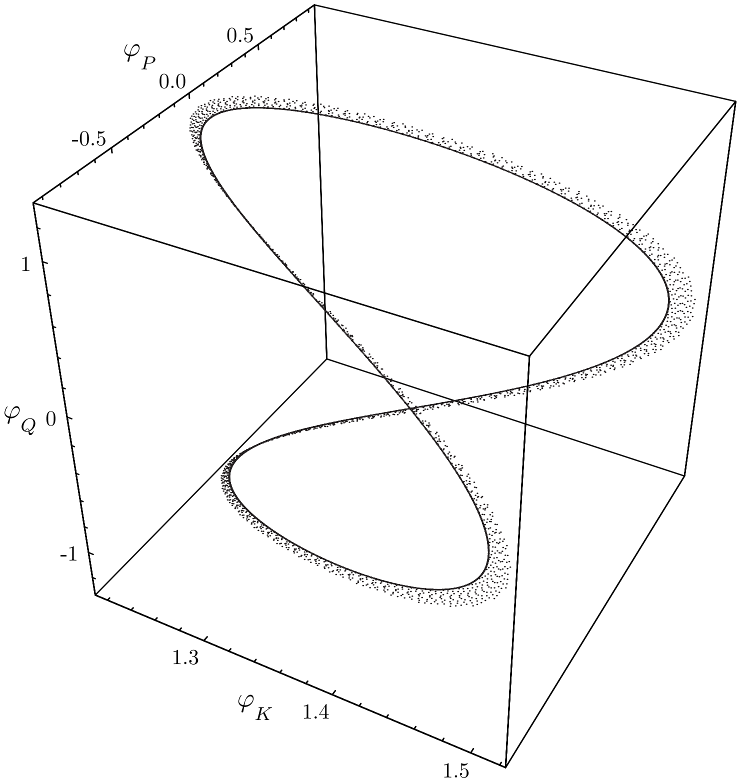

Fig. 3.1. *Integration of the three-wave problem using a conventional second-order predictor–corrector (dotted line) and the conservative predictor–corrector (solid line). Both methods took approximately* 4000 *time steps of size* 0.05*. Initially* $\varphi_K = \sqrt{1.5}$*,* $\varphi_P = 0$*, and* $\varphi_Q = \sqrt{1.5}$*. The effect of the* 4% *energy gain by the conventional method is clearly visible. (From [13].)*

where $\mathfrak{f}_k$ is the discretization of f_k, $\widetilde{S}_k = S_k(\widetilde{\varphi})$ and $\widetilde{\mathfrak{f}}_k = \mathfrak{f}_k(\widetilde{\varphi})$. We expect that it should be sufficient to modify the corrector:

$$\varphi_k^{n+1} = \varphi_k^n + \frac{\tau}{2}\left(S_k(\varphi^n) + \widetilde{S}_k + \mathfrak{g}_k\right). \tag{17}$$

The conservation laws imply

$$\begin{aligned}\frac{\tau}{2}\,\mathfrak{g}_k = &-\left[\varphi_k^n + \frac{\tau}{2}\left(S_k(\varphi^n) + \widetilde{S}_k\right)\right] \\ &+ \operatorname{sgn}(\widetilde{\varphi}_k)\sqrt{(\varphi_k^n)^2 + \tau\left(\varphi_k^n\, S_k(\varphi^n) + \widetilde{\varphi}_k\, \widetilde{S}_k\right)}.\end{aligned} \tag{18}$$

It is straightforward, if somewhat tedious, to show that $\mathfrak{g}_k = \mathcal{O}(\tau^2)$, thus preserving the second-order character of the method. The exactly conservative time-stepper takes the form

$$\begin{aligned}\widetilde{\varphi}_k &= \varphi_k^n + \tau\, S_k(\varphi^n), \\ \varphi_k^{n+1} &= \operatorname{sgn}(\widetilde{\varphi}_k)\sqrt{(\varphi_k^n)^2 + \tau\left(\varphi_k^n\; S_k(\varphi^n) + \widetilde{\varphi}_k\, \widetilde{S}_k\right)}.\end{aligned} \tag{19}$$

Comparing these formulas with the original predictor–corrector we see that original method is altered at *all orders* in τ beyond second.

In Fig. 3.1, we compare the solution of (6) using the conventional predictor–corrector and the conservative predictor–corrector for a large number of orbits. Clearly the conservative method gives superior results. (See [13] for additional examples of conservative integrators and for a more thorough discussion.)

4. Unitary Integrators. As a final example, we consider an n-level quantum system driven by external fields. The dynamics of the density matrix, ρ, generated by a Hamiltonian H, is governed by the quantum Liouville equation:

$$i\hbar\dot{\rho} = [H\,,\rho]\,. \tag{20}$$

This equation has a non-trivial kinematic structure — the Hioe–Eberly invariants, $\operatorname{tr}\rho^j$, $j = 1, 2, \ldots, n$, are non-evolving *regardless* of the form of the Hamiltonian [10]. These constants are a direct consequence of the unitary evolution of the density matrix and are the analogues of the Poincaré invariants in classical mechanics. A numerical solution where these invariants are not preserved is in danger of being unphysical.

The exact dynamics proceeds by unitary evolution

$$\rho(t+T) = \mathcal{U}(t,t+T)\rho(t)\mathcal{U}^\dagger(t,t+T)\,. \tag{21}$$

In the spirit of symplectic integrators discussed in §2, the kinematic structure of (20) will be preserved if the numerical evolution operator is also unitary.

We construct a numerical time-advance mapping by approximating the exact evolution while retaining the unitary property [15]

$$U(t,t+\tau) \equiv e^{-i\tau\,A(t,\tau)} = \mathcal{U}(t,t+\tau) + \mathcal{O}(\tau^\kappa)\,, \tag{22}$$

where A is a Hermitian matrix that will depend on the Hamiltonian.

By matching the Taylor series solution of (20) term-by-term in τ with the approximate evolution generated by $U(t,t+\tau)$, the following approximations for $A(t,\tau)$ are obtained: to second order

$$A = H + \frac{1}{2!}\,\tau\,\dot{H}; \tag{23}$$

to third order

$$A = H + \frac{1}{2!}\,\tau\,\dot{H} + \frac{1}{3!}\,\tau^2\ddot{H} + \frac{i}{12}\,\tau^2\left[H\,,\dot{H}\right]; \tag{24}$$

and to fourth order

$$A = H + \frac{1}{2!}\,\tau\,\dot{H} + \frac{1}{3!}\,\tau^2\ddot{H} + \frac{1}{4!}\,\tau^3\dddot{H} + \frac{i}{12}\,\tau^2\left[H\,,\dot{H}\right] + \frac{i}{4!}\,\tau^3\left[H\,,\ddot{H}\right], \tag{25}$$

where $[\cdot\,,\cdot]$ is the matrix commutator. Note that to obtain accuracy beyond second-order, one must take into account that, in general, $[H(t_1)\,,H(t_2)] \neq 0$.

Fortunately, to use these expressions for U, it is not necessary to exponentiate a (general) $n\times n$ matrix. We are free to approximate A in any way consistent with the order of the method. As described in [14] and [15] it is possible to introduce a suitable basis, $\{\lambda_k\}_{k=1}^{n^2}$, such that exponentials of the basis matrices are easily computed. We then write

$$e^{-i\tau\,A(t,\tau)} = \prod_{k=1}^{n^2} e^{-i\,\tau\,\gamma_k}, \tag{26}$$

where the γ_k are determined by matching terms on each side of (26) order-by-order in τ through $\mathcal{O}(\tau^{\kappa-2})$.

As an example, consider a simple two-level system with a time-independent Hamiltonian

$$H = \begin{pmatrix} \epsilon & \omega \\ \omega & -\epsilon \end{pmatrix} . \tag{27}$$

A second-order integrator is given by

$$\begin{aligned} U(\tau) = & \begin{pmatrix} \cos(\omega\,\tau) & -i\,\sin(\omega\,\tau) \\ -i\,\sin(\omega\,\tau) & \cos(\omega\,\tau) \end{pmatrix} \times \\ & \begin{pmatrix} \cos(\epsilon\,\omega\,\tau^2) & -\sin(\epsilon\,\omega\,\tau^2) \\ \sin(\epsilon\,\omega\,\tau^2) & \cos(\epsilon\,\omega\,\tau^2) \end{pmatrix} \begin{pmatrix} e^{-i\,\epsilon\,\tau} & 0 \\ 0 & e^{i\,\epsilon\,\tau} \end{pmatrix} . \end{aligned} \tag{28}$$

In Fig. 4.1(a) and (b), we show the numerical solution of (20) obtained using the unitary integrator (28) as well as that obtained with a second-order predictor–corrector. The parameters for this example are $\epsilon = 1$, $\omega = 0.01$, and $\tau = 0.1$ and the initial condition is

$$\rho(0) = \frac{1}{2} \begin{pmatrix} 1 & e^{-i\pi/4} \\ e^{i\pi/4} & 1 \end{pmatrix} . \tag{29}$$

In Fig. 4.1(c) and (d) we show the difference between the solutions shown in (a) and (b) and a high-accuracy integration. It is evident that the error of the unitary integrator is much smaller than that of the predictor–corrector even though both methods used the same time step and are both second order. While both methods conserve $\operatorname{tr}\rho$ (it is a linear invariant, and hence preserved by predictor–corrector), this alone is not sufficient to reproduce the dynamics faithfully.

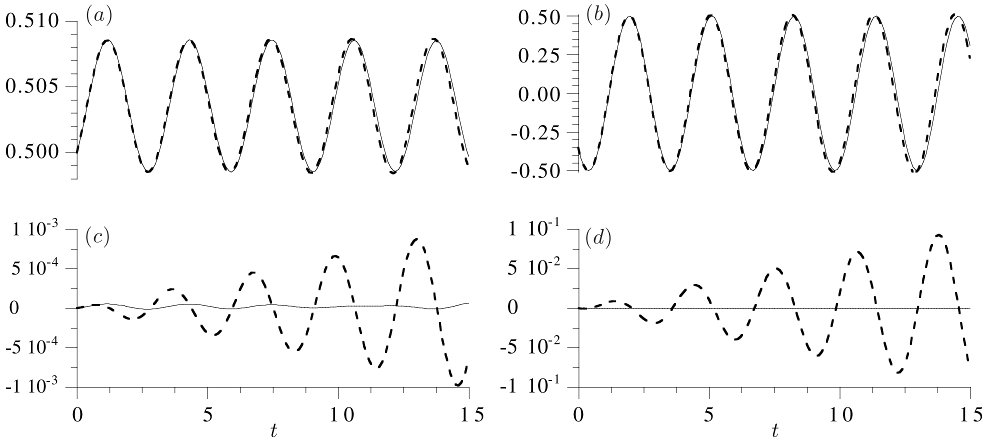

FIG. 4.1. *Results of solving* (20) *using the unitary integrator (solid line) and a second-order predictor–corrector (heavy dashes): numerical solution* ρ_{11} (a) *and* $\operatorname{Im}\rho_{12}$ (b)*; numerical errors in* ρ_{11} (c) *and* $\operatorname{Im}\rho_{12}$ (d).

5. Conclusion. For systems of differential equations that possess structure, better numerical results are obtained when the numerical algorithms used to solve these equations respect this structure. For systems with non-linear algebraic invariants, we have shown that it is possible to systematically construct explicit exactly conservative algorithms. Furthermore, we have seen that kinematic structure is preserved when the numerical time-advance map shares the group properties of the exact dynamics: for Hamiltonian mechanics, we preserve the Poincaré invariants when the time-advance is a canonical transformation; for the quantum Liouville equation, the Hioe–Eberly invariants are preserved when the time-advance is a unitary transformation. Not only are the numerical results superior, but by using structure preserving methods, we are guaranteed the remaining numerical errors do not violate the inherent physics of the models. The key point is that *integrators should have maximum knowledge of the system they are being used to solve.*

REFERENCES

[1] J. A. Armstrong, N. Bloembergen, J. Ducuing, and P. S. Pershan, *Interactions between light waves in a nonlinear dielectric*, Phys. Rev., 127 (1962), pp. 1918–1939.

[2] J. C. Bowman, J. A. Krommes, and M. Ottaviani, *The realizable Markovian closure. I: General theory, with application to three-wave dynamics*, Phys. Fluids B, 5 (1993), pp. 3558–3589.

[3] J. C. Bowman, B. A. Shadwick, and P. J. Morrison, *Exactly conservative integrators*, in 15th IMACS World Congress on Scientific Computation, Modelling and Applied Mathematics, Berlin, August 1997, Wissenschaft & Technik, pp. 595–600.

[4] W. F. Buell and B. A. Shadwick, *Application of unitary integration to interacting and dissipative systems*, Bull. Am. Phys. Soc., 43 (1998), p. 1268.

[5] P. J. Channell and J. C. Scovel, *Symplectic integration of Hamiltonian systems*, Nonlinearity, 3 (1990), pp. 231–259.

[6] R. C. Davidson and A. N. Kaufman, *On the kinetic equation for resonant three-wave coupling*, J. Plasma Phys., 3 (1969), pp. 97–105.

[7] J. de Frutos and J. M. Sanz-Serna, *Erring and being conservative*, in Numerical Analysis 1993, D. F. Griffiths and G. A. Watson, eds., Pitman Research Notes in Mathematics Series, Harlow, 1994, Longmans Scientific and Technical, pp. 75–88.

[8] ———, *Accuracy and conservation properties in numerical integration: the case of the Korteweg–de Vries equation*, Numerische Mathematik, 75 (1997), pp. 421–445.

[9] R. De Vogelære, *Methods of integration which preserve the contact transformation property of the Hamiltonian equations*, Tech. Rep. 4, Department of Mathematics, University of Notre Dame, 1956.

[10] F. T. Hioe and J. H. Eberly, *N-level coherence vector and higher conservation laws in quantum optics and quantum mechanics*, Phys. Rev. Lett., 47 (1981), pp. 838–841.

[11] P. J. Morrison, *Hamiltonian description of the ideal fluid*, Rev. Mod. Phys., 70 (1998), pp. 467–521.

[12] J. M. Sanz-Serna and M. P. Calvo, *Numerical Hamiltonian Problems*, no. 7 in Applied Mathematics and Mathematical Computation, Chapman and Hall, London, 1994.

[13] B. A. Shadwick, J. C. Bowman, and P. J. Morrison, *Exactly conservative integrators*, SIAM J. Appl. Math, 59 (1999).

[14] B. A. Shadwick and W. F. Buell, *Unitary integration with operator splitting for weakly dissipative systems.* In preparation.

[15] ———, *Unitary integration: A numerical technique preserving the structure of the quantum Liouville equation*, Phys. Rev. Lett., 79 (1997), pp. 5189–5193.

[16] G. J. Sussman and J. Wisdom, *Chaotic evolution of the solar system*, Science, 257 (1992), pp. 56–62.

[17] H. Yoshida, *Construction of higher order symplectic integrators*, Phys. Lett. A, 150 (1990), pp. 262–8.

[18] D. M. Young and R. T. Gregory, *A Survey of Numerical Mathematics*, vol. 1, Addison–Wesley, Reading MA, 1972.