

Symplectic and energy conserving discretizations of Hamiltonian PDEs

Jason Frank
Mathematical Institute, Utrecht University

These notes—prepared for guest lectures in the Mastermath course “Numerical methods for time dependent partial differential equations”—address numerical methods for Hamiltonian partial differential equations. Hamiltonian PDEs extend the concept of Hamiltonian ODEs to the infinite dimensional setting, and as such the numerical treatment is largely motivated by results obtained for ODEs. In a sense, our approach is to discretize a Hamiltonian PDE in space so as to obtain a Hamiltonian ODE, for which excellent numerical methods exist. Hamiltonian systems are characterized by their symmetries and conservation laws, and our goal is to preserve such structures when moving from the continuous system to its finite computational analog.

1. A motivating example: point vortices. To motivate these notes, let us consider a very simple model of a fluid such as water or air. If the fluid is incompressible, meaning its density is everywhere uniform, and if it is also inviscid, meaning its internal friction is negligible, then its motion is described by a system of partial differential equations known as the Euler equations, which we will not present here. Instead, we note that when restricted to the plane the Euler equations admit a finite dimensional weak solution comprised of a linear combination of Dirac δ -distributions for the vorticity, a quantity that (roughly speaking) measures the local degree of rotation in the flow. An individual Dirac distribution represents an isolated *point vortex* (e.g. a tornado, hurricane, or something similar).¹ The position at time t of the i th vortex is given by $(X_i(t), Y_i(t))$. In the specific case of two point vortices of equal strength and rotational orientation, the equations of motion are:

$$\begin{aligned} (1) \quad & \dot{X}_1 = (Y_2 - Y_1)[(X_2 - X_1)^2 + (Y_2 - Y_1)^2]^{-1}, \\ (2) \quad & \dot{Y}_1 = (X_1 - X_2)[(X_2 - X_1)^2 + (Y_2 - Y_1)^2]^{-1}, \\ (3) \quad & \dot{X}_2 = (Y_1 - Y_2)[(X_2 - X_1)^2 + (Y_2 - Y_1)^2]^{-1}, \\ (4) \quad & \dot{Y}_2 = (X_2 - X_1)[(X_2 - X_1)^2 + (Y_2 - Y_1)^2]^{-1}. \end{aligned}$$

The above system of differential equations can be solved exactly. However, if we increase the number of point vortices to more than four or five, the motion becomes highly complex and no explicit solutions can be found for generic initial conditions. Instead, the only way to generate solutions is through numerical simulations.

Figure 1 illustrates numerically computed motions of the two vortices as computed using three methods: the forward (or explicit) Euler method, the backward (or implicit) Euler method,

¹In fact, it can be shown that with proper scaling, in the limit as the number of point vortices approaches infinity, the point vortex model weakly approximates smooth solutions of the Euler equations.

and the implicit midpoint rule²:

$$(5) \quad \frac{y_{n+1} - y_n}{\Delta t} = f\left(\frac{y_{n+1} + y_n}{2}\right).$$

The same time step was used for all three methods, namely $\Delta t = 0.02$. With all three methods we observe that the point vortices revolve around each another. This is not surprising: each tornado moves in the wind field induced by the other tornado, which is circular and of equal magnitude. However there are notable differences: in the motion computed with the explicit Euler method the point vortices gradually spiral outward; in the motion computed with the implicit Euler method they repeatedly spiral inward until being ejected; and in the motion computed with the implicit midpoint rule the point vortices revolve without spiraling.

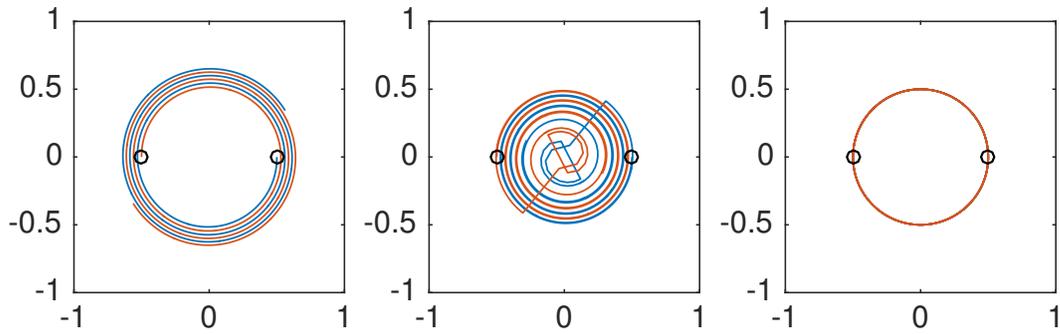


Figure 1. Motion of a pair of point vortices, computed with three numerical methods: forward Euler (left), backward Euler (middle), implicit midpoint rule (right). In all cases, both point vortices have positive orientation and unit circulation, initial positions were $(\pm\frac{1}{2}, 0)$, step size $\Delta t = 0.02$, integration time 625 time steps.

Which motion is correct? Denote the separation distance between the two point vortices by

$$(6) \quad r = [(X_1 - X_2)^2 + (Y_1 - Y_2)^2]^{1/2}.$$

The time rate of change of $r(t)$ along a solution to (1) and (4) is given by

$$\frac{dr}{dt} = \frac{\partial r}{\partial X_1} \dot{X}_1 + \frac{\partial r}{\partial Y_1} \dot{Y}_1 + \frac{\partial r}{\partial X_2} \dot{X}_2 + \frac{\partial r}{\partial Y_2} \dot{Y}_2 = 0,$$

which can be found by substitution. The behavior of the implicit midpoint rule is correct, the spiraling behavior of the Euler methods is not. In Figure 2 on the left we see how the separation distance $r(t)$ changes with time for the three methods. With forward Euler the separation distance grows, with backward Euler it decays until the point vortices “collide” and are dispersed (at which point it decays again), and with implicit midpoint, the separation distance appears constant. On the right of Figure 2 we see that this behavior of the implicit midpoint rule is not just a good approximation: the separation distance is actually

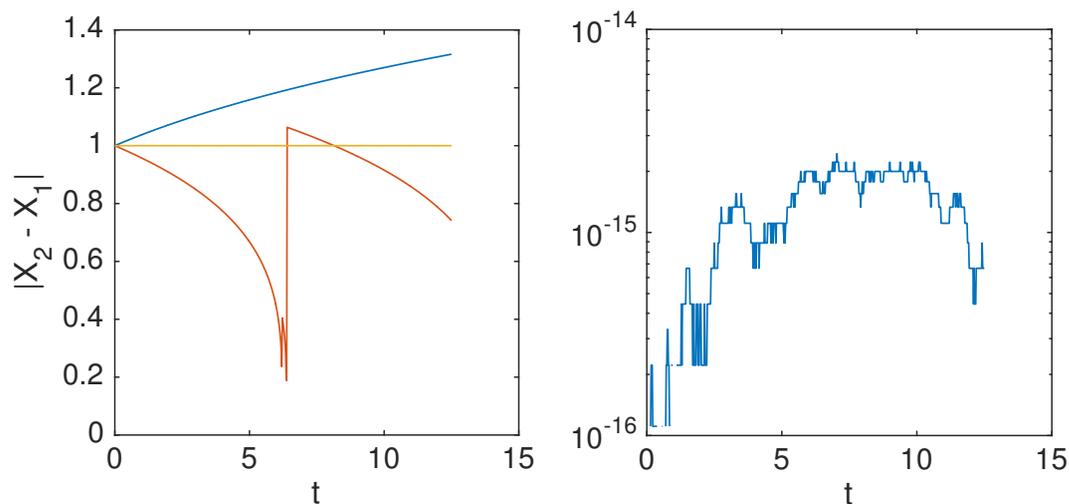


Figure 2. Change in separation $r(t)$ between point vortices as function of time, at left computed with three methods: forward Euler (blue), backward Euler (red), implicit midpoint (yellow). At right, the error $|r(t) - 1|$ as computed with the implicit midpoint rule.

constant to the level of machine precision. We remark that the spiraling behavior of the Euler methods persists if we decrease the stepsize. The rate of spiraling decreases (the methods are convergent, so spiraling must go to zero as $\Delta t \rightarrow 0$), but it is always present for finite step size.

We might be tempted to understand the behavior of these numerical methods via linear stability theory. However we would immediately encounter a problem: linear stability is defined with respect to linearization about an equilibrium, and the system (1)–(4) possesses no equilibria!

On the other hand, the fact that the separation distance r is preserved along any solution implies that all solutions are bounded up to a possible translation of the origin. Furthermore, it may be easily checked that the “center of vorticity”, defined by

$$(7) \quad M = \left(\frac{X_1 + X_2}{2}, \frac{Y_1 + Y_2}{2} \right),$$

also satisfies $dM/dt = 0$ along solutions of (1)–(4). Consequently, we conclude that the two point vortices revolve around a common, fixed center of vorticity, and hence all solutions are bounded.

Summarizing, the adherence to the two conservation laws

$$\frac{dr}{dt} = 0, \quad \frac{dM}{dt} = 0,$$

confines the solutions of (1)–(4) to a compact space, ensuring the boundedness of solutions for all t . The discrete approximate solution obtained with the implicit midpoint rule retains

²Here applied to the generic ODE $y' = f(y)$.

this feature of the continuous solution, whereas neither the solutions computed with forward nor backward Euler does so. In this example we see that conservation is essential to obtain the correct solution behavior on long time intervals. In the first part of these lecture notes we will review the ideas of conservation laws and symmetries in the ODE context and discuss in what ways these can be retained under numerical discretization.

2. Conservation laws. An elementary yet crucial observation from linear algebra is the following. Let $B \in \mathcal{R}^{d \times d}$, $B^T = -B$, be a real skew-symmetric matrix, then

$$(8) \quad x^T B x = 0, \quad \forall x \in \mathcal{R}^d.$$

To check this, define $\alpha = x^T B x$. Note that α is a real scalar quantity, hence equal to its transpose:

$$\alpha = \alpha^T = (x^T B x)^T = x^T B^T x = -(x^T B x) = -\alpha.$$

Since $\alpha = -\alpha$, α must be zero. Skew-symmetry is fundamental to conservation laws.

Consider the autonomous ordinary differential equation

$$(9) \quad \frac{dy}{dt} = f(y), \quad y(t) \in \Lambda \subset \mathcal{R}^d, \quad f : \Lambda \rightarrow \mathcal{R}^d.$$

A function $I : \Lambda \rightarrow \mathcal{R}^d$ that is constant along any solution $y(t)$ of (9), i.e. $I(y(t)) = I(y(0))$, $t > 0$, is called a *first integral*. If $I(y)$ is differentiable, then it follows that for $y(t)$ a solution to (9),

$$(10) \quad 0 = \frac{d}{dt} I(y(t)) = \nabla I(y(t)) \cdot \frac{dy}{dt} = \nabla I(y) \cdot f(y),$$

where $\nabla I(y)$ denotes the gradient of $I(y)$. That is, if I is a first integral, then its gradient is everywhere normal to the vector field $f(y)$. The converse is also true. In physical systems, first integrals represent conservation laws, i.e. quantities that are preserved under the motion of the system.

An obvious class of systems possessing a first integral, are “skew-gradient” systems. Suppose $B(y) : \Lambda \rightarrow \mathcal{R}^{d \times d}$ is a skew-symmetric matrix function, $B(y)^T = -B(y)$, for all $y \in \Lambda$. Then a differential equation of the form

$$(11) \quad \frac{dy}{dt} = B(y) \nabla I(y)$$

is a skew-gradient system. The function $I(y)$ is a first integral of (11), since along a solution,

$$\frac{d}{dt} I(y(t)) = \nabla I(y(t)) \cdot B(y(t)) \nabla I(y(t)) = 0,$$

via the property (8).

Conversely, if $I(y)$ is any first integral of an ODE (9), then away from equilibria y^* of the form $\nabla I(y^*) = 0$, the ODE can be written in the form (11). To see this, think of $f(y)$ and $\nabla I(y)$ as column vectors and define

$$B(y) = \frac{f(y)\nabla I(y)^T - \nabla I(y)f(y)^T}{\nabla I(y)^T \nabla I(y)}.$$

Since $\nabla I(y) \cdot f(y) = 0$, we find

$$\frac{dy}{dt} = B(y)\nabla I(y) = \frac{1}{\|\nabla I(y)\|^2} [f(y)\nabla I(y)^T \nabla I(y) - \nabla I(y)f(y)^T \nabla I(y)] = f(y).$$

We conclude that any system possessing a first integral can be cast in skew-gradient form (11), at least away from equilibria. Skew-gradient structure is in this sense generic (though not necessarily obvious) for systems with conservation laws.

3. Conservation by Runge-Kutta methods. An s -stage Runge-Kutta method for computing an update $y_n \mapsto y_{n+1}$ of the ODE (9) is defined by a set of coefficients (e.g. quadrature weights) b_i , $i = 1, \dots, s$ for the update as well as coefficients a_{ij} , $i, j = 1, \dots, s$ for the s internal stages. The class of Runge-Kutta methods take the form

$$(12) \quad F_i = f \left(y_n + \Delta t \sum_{j=1}^s a_{ij} F_j \right), \quad i = 1, \dots, s,$$

$$(13) \quad y_{n+1} = y_n + \Delta t \sum_{i=1}^s b_i F_i.$$

These methods represent a rather broad class of one step methods. For instance, the forward Euler method is defined by parameter values $s = 1$, $a_{11} = 0$, $b_1 = 1$, backward Euler by parameter values $s = 1$, $a_{11} = 1$, $b_1 = 1$, and the implicit midpoint rule by parameter values $s = 1$, $a_{11} = 1/2$, $b = 1$ (Check this!).

A linear first integral assumes the form $I_1(y) = \kappa^T y + \alpha$, where $\kappa \in \mathcal{R}^d$ is a constant vector and α is a scalar. We encountered two linear first integrals in the example of the point vortices, namely the two components of the center of vorticity vector M defined by (7). It turns out that all linear numerical methods conserve arbitrary linear first integrals. We demonstrate this result for the class of Runge-Kutta methods. Note that the condition (10) becomes $\kappa^T f(y) = 0$, for all $y \in \Lambda$. Multiplying all terms in an s -stage Runge-Kutta method from the left with κ^T yields:

$$\begin{aligned} \kappa^T F_i &= \kappa^T f \left(y_n + \Delta t \sum_{j=1}^s a_{ij} F_j \right) = 0, \quad i = 1, \dots, s, \\ \kappa^T y_{n+1} &= \kappa^T y_n + \Delta t \sum_{i=1}^s b_i \kappa^T F_i = \kappa^T y_n. \end{aligned}$$

It follows that $I_1(y_{n+1}) = \kappa^T y_{n+1} + \alpha = \kappa^T y_n + \alpha = I_1(y_n)$. Hence, $I_1(y_n) = I_1(y_0)$ is constant for all $n > 0$, and we see that Runge-Kutta methods conserve arbitrary linear first integrals (whether or not we are aware of the existence of these).

A quadratic first integral assumes the form $I_2(y) = y^T M y + \kappa^T y + \alpha$, where $M \in \mathcal{R}^{d \times d}$ may, without loss of generality, be assumed to be a symmetric matrix³. The separation distance (6) in the example of the point vortices is a quadratic first integral. One may ask under what, if any, conditions arbitrary first integrals are conserved within the class of Runge-Kutta methods. In this case, the condition (10) implies $2y^T M f(y) + \kappa^T f(y) = 0$. For the forward Euler method, we check

$$\begin{aligned} I_2(y_{n+1}) &= y_{n+1}^T M y_{n+1} + \kappa^T y_{n+1} + \alpha \\ &= [y_n + \Delta t f(y_n)]^T M [y_n + \Delta t f(y_n)] + \kappa^T [y_n + \Delta t f(y_n)] + \alpha \\ &= y_n^T M y_n + \kappa^T y_n + \Delta t [y_n^T M f(y_n) + f(y_n)^T M y_n + \kappa^T f(y_n)] + \Delta t^2 f(y_n)^T M f(y_n) \\ &= I_2(y_n) + \Delta t^2 f(y_n)^T M f(y_n). \end{aligned}$$

For general f , the forward Euler method does not conserve arbitrary quadratic first integrals. How about the implicit midpoint rule (5)? For simplicity, we consider the simpler case $I_2(y) = y^T M y$, which implies $y^T M f(y) = 0$, for all $y \in \Lambda$. Define $\hat{y} = (y_{n+1} + y_n)/2$. Multiplying both sides of (5) from the left by $\hat{y}^T M$ yields

$$\left(\frac{y_{n+1} + y_n}{2} \right)^T M \left(\frac{y_{n+1} - y_n}{\Delta t} \right) = \hat{y}^T M f(\hat{y}) = 0.$$

Simplifying the first term of this relation yields $I_2(y_{n+1}) - I_2(y_n) = 0$.

More generally it is straightforward to check that any Runge-Kutta method whose coefficients satisfy the condition:

$$(14) \quad b_i b_j - b_i a_{ij} - b_j a_{ji} = 0, \quad i, j = 1, \dots, s$$

conserves arbitrary quadratic first integrals. (Check this! Note that (10) implies the relation $\kappa^T f(y) = -2y^T M f(y)$, $\forall y \in \Lambda$. Check that implicit midpoint satisfies this criterion.) Higher order generalizations of the implicit midpoint rule are called *Gauss-Legendre collocation methods*.

How about other polynomial first integrals? Unfortunately we are out of luck. There are no Runge-Kutta methods that automatically preserve all cubic first integrals, for instance.

To preserve other first integrals, we must design tailored methods for specific conservation laws. One approach to doing so utilizes the skew-gradient form (11) and a so-called “discrete gradient”.

To construct a discrete gradient method, we need to define a matrix-valued function \bar{B} and a vector-valued function $\bar{\nabla} I$, each taking *two* arguments. Let $\bar{B}(y, z) : \mathcal{R}^d \times \mathcal{R}^d \rightarrow \mathcal{R}^{d \times d}$ satisfy the following conditions:

³The skew-symmetric part of M contributes nothing to the value of the integral due to (8).

- skew-symmetric structure: $\bar{B}(y, z)^T = -\bar{B}(y, z)$,
- symmetry with respect to its arguments: $\bar{B}(y, z) = \bar{B}(z, y)$,
- consistency $\bar{B}(y, y) = B(y)$.

Similarly define a function $\bar{\nabla}I(y, z) : \mathcal{R}^d \times \mathcal{R}^d \rightarrow \mathcal{R}^d$ that satisfies the condition

$$\bar{\nabla}I(y, z)^T(y - z) = I(y) - I(z),$$

as well as the consistency condition $\bar{\nabla}I(y, y) = \nabla I(y)$. Subsequently, the method

$$y_{n+1} = y_n + \tau \bar{B}(y_{n+1}, y_n) \bar{\nabla}I(y_{n+1}, y_n)$$

satisfies $I(y_{n+1}) = I(y_n)$, since

$$\begin{aligned} I(y_{n+1}) - I(y_n) &= \bar{\nabla}I(y_{n+1}, y_n)^T(y_{n+1} - y_n) \\ &= \tau \bar{\nabla}I(y_{n+1}, y_n)^T \bar{B}(y_{n+1}, y_n) \bar{\nabla}I(y_{n+1}, y_n) = 0. \end{aligned}$$

One possible discrete gradient is the *midpoint discrete gradient*. Denote $y_{n+1/2} := (y_{n+1} + y_n)/2$ and $\Delta y := y_{n+1} - y_n$. Then

$$\bar{\nabla}I(y_{n+1}, y_n) := \nabla I(y_{n+1/2}) + \frac{I(y_{n+1}) + I(y_n) - \nabla I(y_{n+1/2})^T \Delta y}{\|\Delta y\|^2} \Delta y.$$

A possible realization of the matrix \bar{B} is given by

$$\bar{B}(y, z) = B\left(\frac{y+z}{2}\right).$$

Discrete gradient methods are always implicit. With a little creativity, however, one can construct splitting methods that are a composition of discrete gradient methods each of which is implicit in only one variable (i.e. requiring only a scalar nonlinear solve for each flow).

4. Hamiltonian systems. A Hamiltonian system is a differential equation on \mathcal{R}^{2d} with an associated distinguished function $H(q, p) : \mathcal{R}^d \times \mathcal{R}^d \rightarrow \mathcal{R}$ called the *Hamiltonian*, such that the differential equation is written

$$(15) \quad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad i = 1, \dots, d.$$

The structure of these equations makes it clear that the Hamiltonian is conserved along a solution $(q(t), p(t))$, since

$$(16) \quad \frac{d}{dt}H(q(t), p(t)) = \sum_{i=1}^d \frac{\partial H}{\partial q_i} \frac{dq_i}{dt} + \frac{\partial H}{\partial p_i} \frac{dp_i}{dt} = \sum_{i=1}^d \frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} = 0,$$

where we have substituted the equations of motion (15). In most applications, H represents the total energy, and (15) are the equations of motion of an energy conserving (or conservative) system.

Above we have already tacitly introduced the vector notation $q(t) = (q_1(t), \dots, q_d(t))^T$, $p(t) = (p_1(t), \dots, p_d(t))^T$. We also introduce the notation

$$\frac{\partial H}{\partial q} = \left(\frac{\partial H}{\partial q_1}, \dots, \frac{\partial H}{\partial q_d} \right)^T, \quad \frac{\partial H}{\partial p} = \left(\frac{\partial H}{\partial p_1}, \dots, \frac{\partial H}{\partial p_d} \right)^T,$$

and write (15) as

$$\frac{dq}{dt} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial q}.$$

Yet more generally, we may work with a vector $y(t) \in \mathcal{R}^d$, $H(y) : \mathcal{R}^d \rightarrow \mathcal{R}$, and write a generic Hamiltonian system in skew-gradient form

$$(17) \quad \frac{dy}{dt} = J^{-1} \frac{\partial H}{\partial y},$$

where $J = -J^T$ is a constant, invertible, skew-symmetric matrix. In other words, the subclass of skew-gradient systems with $B(y) = J = \text{const}$ are the generalized Hamiltonian systems. The particular case $y(t) \in \mathcal{R}^{2d}$ and

$$J^{-1} = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$

(where I is the identity matrix on \mathcal{R}^d) corresponds to (15) and is referred to as a *canonical Hamiltonian system*.

Example. The point vortex system (1)–(2) is a canonical Hamiltonian system with $y = (X_1, X_2, Y_1, Y_2)^T$ and Hamiltonian function

$$H = \log r$$

More generally, a point vortex system with N point vortices with *vortex strengths* $\Gamma_i \in \mathcal{R}$, $i = 1, \dots, N$, is given by

$$(18) \quad \Gamma_i \dot{X}_i = \frac{\partial H}{\partial Y_i}, \quad \Gamma_i \dot{Y}_i = -\frac{\partial H}{\partial X_i}, \quad H = \sum_{i=1}^N \sum_{j=i+1}^N \Gamma_i \Gamma_j \log [(X_i - X_j)^2 + (Y_i - Y_j)^2].$$

5. Symplectic maps. Given a Hamiltonian system the skew-symmetric matrix J can be used to define a bilinear form

$$\omega(a, b) = a^T J b, \quad \forall a, b \in \mathcal{R}^d.$$

This form plays an important role in symplectic geometry, but we will not go into that here. Instead, we call a matrix A a *symplectic matrix* if A defines a linear transformation that leaves ω invariant in the sense that

$$\omega(Aa, Ab) = \omega(a, b), \quad \forall a, b \in \mathcal{R}^d.$$

Since this has to hold for all vectors a and b , we can choose the canonical vectors $e_i, i = 1, \dots, d$ to prove that A is symplectic if and only if

$$A^T J A = J.$$

Similarly, a map $g(y) : \mathcal{R}^d \rightarrow \mathcal{R}^d$ is said to be a *symplectic transformation* if its Jacobian $Dg(y)$ satisfies

$$Dg(y)^T J Dg(y) = J.$$

The fundamental property of a Hamiltonian system is that its flow is a symplectic transformation. The flow $\Phi_t : \Lambda \rightarrow \Lambda$ is a one-parameter map. For a given $t \in \mathcal{R}$, it maps an initial condition y_0 to the solution of (17) at time t . For a given y_0 , the solution $y(t)$ of (17) satisfies

$$y(t) = \Phi_t(y_0), \quad \forall t \in \mathcal{R}.$$

We demonstrate that the Jacobian $D\Phi_t$ is a symplectic map, i.e. satisfies

$$(D\Phi_t)^T J D\Phi_t = J.$$

To see that this holds, define the matrix function $M(t) = D\Phi_t(y_0)$, and note that this matrix function satisfies the fundamental matrix equation:

$$\dot{M}(t) = Df(\Phi_t(y_0))M(t)$$

In the case of a Hamiltonian system, we have

$$f(y) = J^{-1}\nabla H(y) \quad \Rightarrow \quad Df(y) = J^{-1}H_{yy},$$

with H_{yy} the (symmetric) Hessian matrix of second derivatives $\frac{\partial^2 H}{\partial y_i \partial y_j}$. Hence $M(t)$ solves

$$\dot{M} = J^{-1}H_{yy}M.$$

Next, differentiate the matrix $R = M^T J M$ with respect to time:

$$\begin{aligned} \dot{R} &= \dot{M}^T J M + M^T J \dot{M}, \\ \dot{R} &= M^T H_{yy} (J^{-1})^T J M + M^T J J^{-1} H_{yy} M. \end{aligned}$$

Noting that $(J^{-1})^T = -J^{-1}$, we find

$$\dot{R} = -M^T H_{yy} M + M^T H_{yy} M = 0.$$

Furthermore, since $M(0) = \Phi'_0(y) = I$, we must have

$$R(0) = M(0)^T J M(0) = J.$$

Since the matrix R is constant in time, we must have the desired identity $M^T J M = J$.

6. Symplectic integrators. Analogous to the flow Φ_t of a differential equation, we can define the map associated to a numerical method $\Psi_{\Delta t}$, i.e. $y_{n+1} = \Psi_{\Delta t}(y_n)$. For example, the forward Euler map associated with the ODE (9) is

$$\Psi_{\Delta t}(y) = y + \Delta t f(y).$$

A numerical method is called a *symplectic method* if its map is symplectic, i.e.

$$D\Psi_{\Delta t}(y)^T J D\Psi_{\Delta t}(y) = J, \quad \forall y \in \Lambda.$$

One can check that the implicit midpoint rule is a symplectic method. In fact, a Runge-Kutta method is symplectic if and only if⁴ it satisfies the condition (14) for conserving arbitrary quadratic first integrals⁵.

Besides Runge-Kutta methods (and the related important class of *partitioned Runge-Kutta methods*), another (often more attractive) class of symplectic integrators can be derived by splitting.

Splitting methods are based on two facts: (1) the exact flow of a Hamiltonian system is symplectic, as shown in the previous section; and (2) the composition of symplectic maps is again symplectic. To see the second of these, note that if f and g are both symplectic maps on Λ with corresponding Jacobians Df and Dg , then the transformation $f(g(y))$ satisfies:

$$\begin{aligned} D[f(g(y))]^T J D[f(g(y))] &= [Df(g(y))Dg(y)]^T J Df(g(y))Dg(y) \\ &= Dg(y)^T Df(g)^T J Df(g(y))Dg(y) \\ &= Dg(y)^T J Dg(y) \\ &= J. \end{aligned}$$

Now all we have to do to construct a symplectic splitting method is to split the Hamiltonian $H(y)$ into a number, say K , of terms

$$H(y) = H_1(y) + H_2(y) + \cdots + H_K(y)$$

such that for each k the motion

$$\dot{y} = J^{-1} \nabla H_k(y)$$

can be exactly solved analytically to obtain an elementary flow Φ_t^k . Then we construct a numerical method by composing these flows over a time step Δt :

$$\Psi_{\Delta t} = \Phi_{\Delta t}^K \circ \cdots \circ \Phi_{\Delta t}^1.$$

Each elementary flow is symplectic because it is the exact flow of a Hamiltonian system, and the composite method is symplectic because it is a composition of symplectic flows. The

⁴Strictly speaking, the ‘only if’ also requires the method to be irreducible.

⁵This is all related to the quadratic form of the conservation law for R in the previous section, but that discussion would take us too far astray here.

above method is first order accurate. A second order variant can be obtained by composing the above flow with its adjoint

$$\Psi_{\Delta t} = \Phi_{\Delta t/2}^1 \circ \cdots \circ \Phi_{\Delta t/2}^{K-1} \circ \Phi_{\Delta t}^K \circ \Phi_{\Delta t/2}^{K-1} \circ \cdots \circ \Phi_{\Delta t/2}^1.$$

(Note that the innermost two flows are identical and commute, hence can be combined into one.)

It might seem difficult to find a suitable splitting such that the elementary flows are all exactly solvable. Experience shows however, that this is very often possible. Some particular cases of solvable flows are linear flows, which can be integrated using the matrix exponential function or the ‘separation of variables’ formula; and shear flows in which one variable remains constant implying the derivative of another one remains constant. This last case occurs often in classical mechanical systems, where the Hamiltonian is separable, i.e.

$$H(q, p) = \frac{1}{2}p^T p + V(q), \quad q, p \in \mathcal{R}^d, \quad V : \mathcal{R}^d \rightarrow \mathcal{R}^d.$$

In this case we can just take $H_1 = \frac{1}{2}p^T p$ and $H_2 = V(q)$. The flow generated by H_1 is

$$\dot{q} = p, \quad \dot{p} = 0,$$

with exact solution $q(t + \Delta t) = q(t) + p(t)$, $p(t + \Delta t) = p(t)$. The flow generated by H_2 is

$$\dot{q} = 0, \quad \dot{p} = -\nabla V(q),$$

with exact solution $q(t + \Delta t) = q(t)$, $p(t + \Delta t) = p(t) - \Delta t \nabla V(q(t))$.

The second order, self-adjoint composition of these two flows leads to the *Störmer-Verlet* method

$$\begin{aligned} q_{n+1/2} &= q_n + \frac{\Delta t}{2} p_n, \\ p_{n+1} &= p_n - \Delta t \nabla V(q_{n+1/2}), \\ q_{n+1} &= q_{n+1/2} + \frac{\Delta t}{2} p_{n+1} \end{aligned}$$

This method is second order, self-adjoint, symplectic, and explicit! When V is a complicated nonlinear function, its gradient need only be evaluated once per time step. This method is heavily used in molecular dynamics. It does *not* fit the class of standard Runge-Kutta methods.

Example. One can construct a symplectic splitting method for the N -point vortex system (18) by splitting the Hamiltonian into *pair interactions*

$$H = \sum_{i=1}^N \sum_{j=i+1}^N H_{ij}, \quad H_{ij} = \Gamma_i \Gamma_j \log [(X_i - X_j)^2 + (Y_i - Y_j)^2]$$

The flow generated by one of the H_{ij} is just a motion of two point vortices with all other point vortices held fixed. The separation distance (6) is conserved under this motion (it is a function of the Hamiltonian H_{ij}). Additionally, another conservation law will hold. For instance, suppose for $i = 1$ and $j = 2$ we have $\Gamma_i = \Gamma_j = 1$. Then the pair interaction becomes (1)–(4). The Hamiltonian H_{ij} is constant, implying that r is constant, and therefore the denominator in (1)–(4) is constant. The equations reduce to

$$\dot{X}_1 = \frac{Y_2 - Y_1}{r(0)^2}, \quad \dot{Y}_1 = \frac{X_1 - X_2}{r(0)^2}, \quad \dot{X}_2 = \frac{Y_1 - Y_2}{r(0)^2}, \quad \dot{Y}_2 = \frac{X_2 - X_1}{r(0)^2}.$$

This system is linear and can be solved either as a series of rotations, or by applying the matrix exponential. The whole splitting method then involves looping over all pairs, calculating their mutual distance, computing a matrix exponential, and updating that pair.

7. Backward error analysis. Symplectic numerical methods share the symplectic property with the exact flow of Hamiltonian systems. So what? Does this have any benefits for computation? Yes it does. Numerical experiments with symplectic methods indicate excellent (though not exact) conservation of the Hamiltonian as well as other first integrals when the system is integrable.

To understand this, we need to apply backward error analysis. To do so, we derive a modified differential equation that better “approximates” the numerical flow $\Psi_{\Delta t}$. We illustrate this idea for the forward Euler method applied to the ODE (9). We look for a modified differential equation as a correction to (9):

$$\dot{y} = f(y) + \Delta t f_1(y)$$

The exact solution to this modified ODE at time $t + \Delta t$ is given by Taylor expansion

$$\begin{aligned} y(t + \Delta t) &= y(t) + \Delta t \dot{y} + \frac{\Delta t^2}{2} \ddot{y} + \dots \\ &= y(t) + \Delta t f(y(t)) + \Delta t^2 \left[f_1(y(t)) + \frac{1}{2} Df(y(t)) f(y(t)) \right] + \mathcal{O}(\Delta t^3) \end{aligned}$$

Now the idea is to choose the vector field $f_1(y)$ such that the forward Euler map when applied to (9),

$$y_{n+1} = y_n + \Delta t f(y_n),$$

agrees with the above solution up to $\mathcal{O}(\Delta t^3)$. Clearly this will be the case for $f_1(y) = -\frac{1}{2} Df(y) f(y)$, leading to the modified equation

$$(19) \quad \dot{y} = f(y) - \frac{\Delta t}{2} Df(y) f(y).$$

Note the following:

- Whereas forward Euler has an $\mathcal{O}(\Delta t^2)$ error with respect to the solution of the original ODE (9), forward Euler applied to (9) agrees with the solution of (19) to $\mathcal{O}(\Delta t^3)$. In this sense, the exact solution of (19) better describes the numerical solution of Euler than does (9).

- The modified equation (19) explicitly depends on the time step Δt used to integrate the original ODE with forward Euler. As $\Delta t \rightarrow 0$, the modified equation converges to the original ODE, as the numerical method converges to the solution.
- We have derived the lowest order correction of the modified equation. We could continue, finding a new modified equation $\dot{y} = f(y) + \Delta t f_1(y) + \Delta t^2 f_2(y)$, and carrying out the Taylor expansion to $\mathcal{O}(\Delta t^4)$ to find an even more accurate modified equation. In fact, we continue this as much as we please, expressing the modified equation as an asymptotic expansion. For any fixed Δt , such an expansion fails to converge but may be optimally truncated.

Now comes the crux. If we carry out a backward error analysis as above for a symplectic method of order p , we find the remarkable result that the modified equation takes the form

$$\dot{y} = J^{-1} \nabla \tilde{H}(y), \quad \tilde{H}(y) = H(y) + \Delta t^p H_p(y) + \Delta t^{p+1} H_{p+1}(y) + \dots$$

In other words. The modified equation is *again a Hamiltonian system* with a perturbed Hamiltonian function. The modified Hamiltonian \tilde{H} is a first integral of the modified equation, and hence it is conserved by the method, i.e. $\tilde{H}(y_{n+1}) = \tilde{H}(y_n)$. However, since

$$H(y) = \tilde{H}(y) + \mathcal{O}(\Delta t^p),$$

we observe that the original Hamiltonian is approximately conserved up to fluctuations with an amplitude of $\mathcal{O}(\Delta t^p)$.

Example. As an example consider the nonlinear pendulum. Let q be the angle, measured clockwise from the vertical at the lowest point, and let $p = \dot{q}$. The equations of motion are

$$\begin{aligned} \dot{q} &= p \\ \dot{p} &= -\sin q. \end{aligned}$$

And the Hamiltonian is

$$(20) \quad H = \frac{p^2}{2} - \cos q.$$

We will compare three methods for this problem, forward Euler, backward Euler, and the following method, which we refer to as *symplectic Euler*

$$\begin{aligned} q_{n+1} &= q_n + \Delta t p_n \\ p_{n+1} &= p_n - \Delta t \sin q_{n+1}. \end{aligned}$$

Note that this method can be computed explicitly even though q is evaluated at time level $n + 1$ on the right. The method is first order accurate.

The solutions obtained with all three methods are plotted in Figure [Figure 3](#). The black curves are level curves of the energy (20). Most solutions are periodic; the pendulum either swings

back and forth or around and around. There is a center equilibrium point at the bottom of its arc and a saddle point at the top.

The numerical solutions with all three methods were computed using step size $\Delta t = 0.1$ for 200 steps, with initial condition $q(0) = -3\pi/4$, $p(0) = 0$. The exact solution through this point is a closed energy level set, the pendulum swings back and forth. The forward Euler solution is seen to spiral outwards, gaining speed until it crosses the separatrix. The backward Euler solution spirals inward towards the center equilibrium point. The symplectic Euler solution appears to be a closed periodic orbit. What does backward error analysis tell us about this problem?

The Jacobian for the pendulum is

$$f'(q, p) = \begin{bmatrix} 0 & 1 \\ -\cos q & 0 \end{bmatrix}.$$

The $\mathcal{O}(\Delta t)$ perturbation in the modified vector field for the forward Euler method is

$$f_1 = -\frac{1}{2}f'f = \begin{pmatrix} -\frac{1}{2}\sin q \\ -\frac{1}{2}p \cos q \end{pmatrix}.$$

This vector field is plotted in the leftmost frame of Figure [Figure 4](#). Note that the component of the vector field that is normal to the level sets of constant energy is everywhere pointing outwards. The center point becomes an unstable equilibrium. The effect of this perturbation is an increase in energy along each trajectory, which is consistent with the behavior exhibited by forward Euler.

It may come as no surprise that the $\mathcal{O}(\Delta t)$ perturbation of backward Euler is just the negative of that of forward Euler, as was also the case for the principle term in the local error of backward Euler. We have

$$f_1 = \frac{1}{2}f'f = \begin{pmatrix} \frac{1}{2}\sin q \\ \frac{1}{2}p \cos q \end{pmatrix},$$

and this vector field is plotted in the second frame of Figure [Figure 4](#). In this case the perturbation causes a decrease in energy along each trajectory, and the center point becomes a stable equilibrium.

For the symplectic method, we expand the numerical solution for p_{n+1} about time t_n

$$\begin{aligned} q_{n+1} &= q_n + \Delta t p_n \\ p_{n+1} &= p_n - \Delta t(\sin q_n + \Delta t \dot{q}_n \cos q_n + \mathcal{O}(\Delta t^2)) \\ &= p_n - \Delta t \sin q_n - \Delta t^2 p_n \cos q_n + \mathcal{O}(\Delta t^3). \end{aligned}$$

The term f_1 must to be chosen such that the $\mathcal{O}(\Delta t^2)$ terms agree:

$$f_1(q_n, p_n) + \frac{1}{2}f'(q_n, p_n)f(q_n, p_n) = \begin{pmatrix} 0 \\ -p_n \cos q_n \end{pmatrix},$$

that is,

$$f_1(q, p) = \begin{pmatrix} \frac{1}{2} \sin q_n \\ -\frac{1}{2} p_n \cos q_n \end{pmatrix}.$$

The pendulum equations can be written in the form of a skew-symmetric matrix times a gradient, as in the previous section, to make it more explicit that (20) is conserved. Let $y = (q, p)^T$; then

$$y' = J^{-1} \nabla H(y), \quad J^{-1} := \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

The modified vector field f_1 can also be written as a product of J^{-1} and a gradient, namely,

$$f_1 = J^{-1} \nabla H_1(y), \quad H_1(q, p) = \frac{1}{2} p \sin q$$

It follows that the 1-term modified equation is

$$\tilde{y}' = J \nabla (H(\tilde{y}) + \Delta t H_1(\tilde{y})),$$

which also has a conserved quantity $\tilde{H} = H + \Delta t H_1$. In particular, this means that the 1-term modified equation also has periodic solutions. The level sets of \tilde{H} are plotted in the third frame of Figure 4, using a larger stepsize perturbation $\Delta t = 0.2$ to exaggerate the effect.

Note that \tilde{H} is *not* a conserved quantity of the numerical method, since we have only used a single term in the modified equation expansion. However, since the symplectic Euler method is a second order accurate approximation to this modified equation, we would expect that \tilde{H} to be better preserved than H .

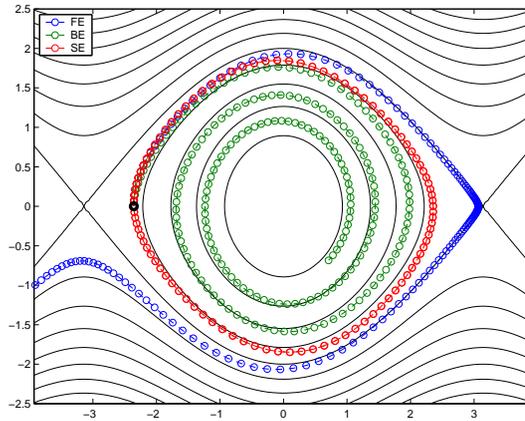


Figure 3. Numerical solutions of the nonlinear pendulum superimposed over the energy level sets.

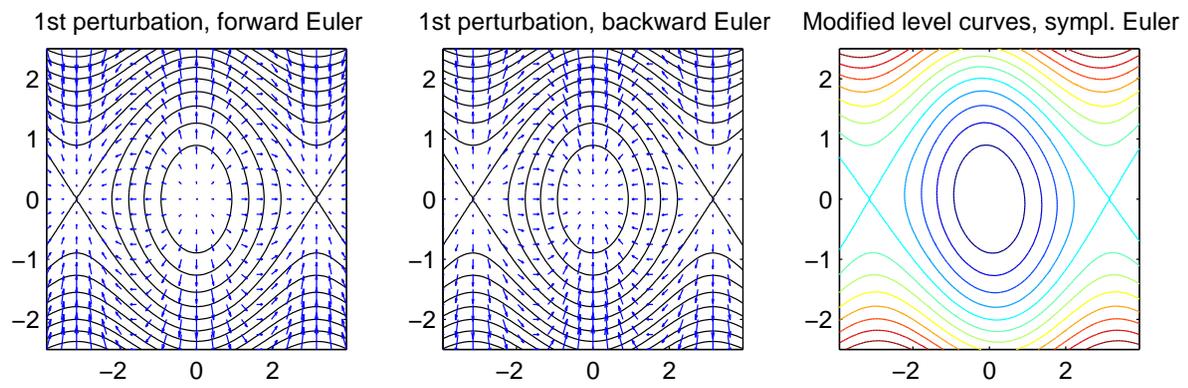


Figure 4. Vector field perturbation f_1 for forward Euler (left) and backward Euler (center). On the right are the energy level sets of the 1-term modified equation for the symplectic Euler method.

8. The Korteweg-de Vries equation. A *soliton* was famously first observed in nature by civil engineer John Scott Russell in 1834, while studying wave motion for the design of canal boats. He wrote:

I was observing the motion of a boat which was rapidly drawn along a narrow channel by a pair of horses, when the boat suddenly stopped—not so the mass of water in the channel which it had put in motion; it accumulated round the prow of the vessel in a state of violent agitation, then suddenly leaving it behind, rolled forward with great velocity, assuming the form of a large solitary elevation, a rounded, smooth and well-defined heap of water, which continued its course along the channel apparently without change of form or diminution of speed. I followed it on horseback, and overtook it still rolling on at a rate of some eight or nine miles an hour [14 km/h], preserving its original figure some thirty feet [9 m] long and a foot to a foot and a half [300–450 mm] in height. Its height gradually diminished, and after a chase of one or two miles [2–3 km] I lost it in the windings of the channel. Such, in the month of August 1834, was my first chance interview with that singular and beautiful phenomenon which I have called the Wave of Translation. (*Source: Wikipedia*)

A soliton is a localized nonlinear wave form that translates at constant speed without deforming. The Korteweg-de Vries equation

$$(21) \quad u_t - 6uu_x + u_{xxx} = 0$$

is an example of a nonlinear wave equation that supports soliton solutions of the form

$$u(x, t) = -\frac{c}{2} \operatorname{sech}^2 \left[\frac{\sqrt{c}}{2}(x - ct) \right],$$

where c is the soliton wave speed. The Korteweg-de Vries equation is an example of a Hamiltonian partial differential equation.

9. Variational derivatives. Hamiltonian partial differential equations constitute an analog of the above concepts (i.e. Hamiltonian ODEs) on an infinite dimensional phase space. Their proper definition requires a few more formal notions.

First, we define a connected domain $\mathcal{D} \subset \mathcal{R}$. We generally choose the interval $\mathcal{D} = [0, L]$. Let \mathcal{V} denote a space of smooth functions $\mathcal{D} \rightarrow \mathcal{R}^d$. Note these are vector valued functions of space and time in general. We do not specify \mathcal{V} but will assume that its elements are smooth with smooth derivatives to a certain order as needed for our application. We also assume that \mathcal{V} is equipped with an inner product denoted $\langle \cdot, \cdot \rangle$ such that

$$\langle u, v \rangle = \int_0^L u(x)^T v(x) dx, \quad \forall u, v \in \mathcal{V}.$$

We will define Hamiltonian partial differential equations on \mathcal{V} . For the purposes of these notes we assume periodic boundary conditions $u(x + L) = u(x)$ for all $u \in \mathcal{V}$.

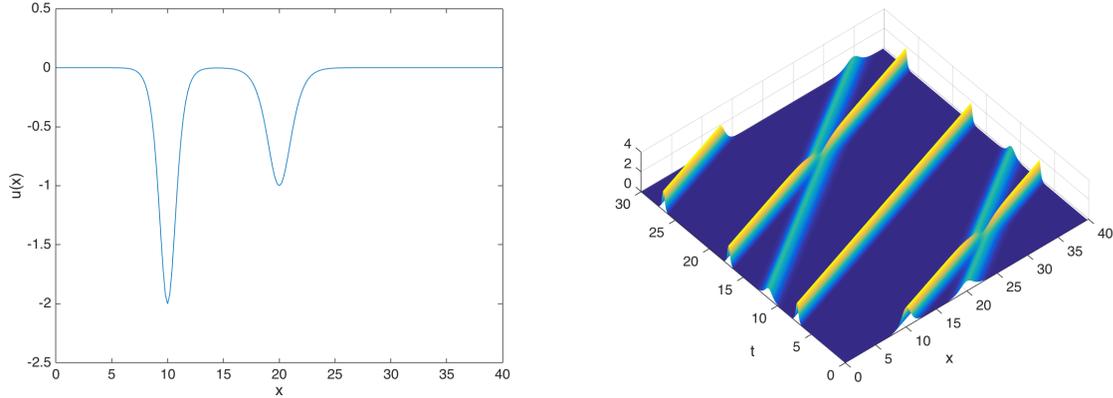


Figure 5. A collision of two solitons in the KdV equation. Left: initial condition; right: space-time plot of $-u(x,t)$.

A functional $\mathcal{F}[u] : \mathcal{V} \rightarrow \mathcal{R}$ is defined as an integral of functions of u and its spatial derivatives u_x, u_{xx} , and so forth. For example, we may write

$$\mathcal{F}[u] = \int_0^L f(u, u_x, u_{xx}) dx = \int_0^L u^3 + \frac{u_x^2}{2} dx.$$

The notation $\mathcal{F}[u]$ emphasizes the functional dependence of \mathcal{F} on its argument u .

We may define a derivative of the functional \mathcal{F} with respect to its argument, the function u . Such a derivative is called a *variational derivative*. To do so we go back to the definition of the derivative. The variational derivative of \mathcal{F} , denoted $\frac{\delta \mathcal{F}}{\delta u}$ is defined weakly through the relation

$$(22) \quad \left\langle \frac{\delta \mathcal{F}}{\delta u}, v \right\rangle = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (\mathcal{F}[u + \varepsilon v] - \mathcal{F}[u]), \quad \forall v \in \mathcal{V}.$$

Let us consider some examples. Consider the functional

$$\mathcal{F}[u] = \int_0^L u^3 dx.$$

where $u(x) : \mathcal{D} \rightarrow \mathcal{R}$ is scalar valued. Note that

$$\mathcal{F}[u + \varepsilon v] = \int_0^L (u + \varepsilon v)^3 dx.$$

The variational derivative satisfies

$$\begin{aligned}
\left\langle \frac{\delta \mathcal{F}}{\delta u}, v \right\rangle &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_0^L (u + \varepsilon v)^3 - u^3 dx, \\
&= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_0^L u^3 + 3\varepsilon u^2 v + 3\varepsilon^2 u v^2 + \varepsilon^3 v^3 - u^3 dx, \\
&= \lim_{\varepsilon \rightarrow 0} \int_0^L 3u^2 v + \varepsilon 3u v^2 + \varepsilon^2 v^3 dx, \\
&= \int_0^L 3u^2 v dx, \\
&= \langle 3u^2, v \rangle.
\end{aligned}$$

Therefore we have $\frac{\delta \mathcal{F}}{\delta u} = 3u^2$. Note that in this case, the variational derivative is just the usual derivative of the integrand u^3 .

As a second example, consider the functional

$$\mathcal{F}[u] = \int_0^L \frac{u_x^2}{2} dx.$$

In this case we have

$$\mathcal{F}[u + \varepsilon v] = \int_0^L \frac{(u_x + \varepsilon v_x)^2}{2} dx.$$

We compute the variational derivative

$$\begin{aligned}
\left\langle \frac{\delta \mathcal{F}}{\delta u}, v \right\rangle &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_0^L \frac{(u_x + \varepsilon v_x)^2}{2} - \frac{u_x^2}{2} dx, \\
&= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_0^L \frac{1}{2} (u_x^2 + 2\varepsilon u_x v_x + \varepsilon^2 v_x^2 - u_x^2) dx, \\
&= \lim_{\varepsilon \rightarrow 0} \int_0^L u_x v_x + \frac{\varepsilon}{2} v_x^2 dx, \\
&= \int_0^L u_x v_x dx.
\end{aligned}$$

This relation is not yet in the form of (22) because we cannot write it as an inner product with the (arbitrary) function v . To do so, we need to integrate by parts, and this requires using the boundary conditions. We compute

$$\int_0^L u_x v_x dx = \int_0^L \frac{\partial}{\partial x} (u_x v) - u_{xx} v dx = u_x v \Big|_0^L + \int_0^L (-u_{xx}) v dx.$$

Since we work with periodic boundary conditions, the first term is zero. The second term is in the correct form:

$$\left\langle \frac{\delta \mathcal{F}}{\delta u}, v \right\rangle = \langle -u_{xx}, v \rangle.$$

From this, we see that $\frac{\delta \mathcal{F}}{\delta u} = -u_{xx}$.

One may check that for a functional of the form

$$\mathcal{F}[u] = \int_0^L f(u, u_x, u_{xx}, u_{xxx}, \dots) dx,$$

the variational derivative is

$$\frac{\delta \mathcal{F}}{\delta u} = \frac{\partial f}{\partial u} - \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial u_x} \right) + \frac{\partial^2}{\partial x^2} \left(\frac{\partial f}{\partial u_{xx}} \right) - \frac{\partial^3}{\partial x^3} \left(\frac{\partial f}{\partial u_{xxx}} \right) + \dots$$

For example, consider

$$\mathcal{F}[u] = \int_0^L uu_x^2 dx.$$

Here we have $f(u, u_x) = uu_x^2$, $\partial f / \partial u = u_x^2$, $\partial f / \partial u_x = 2uu_x$. The variational derivative is

$$\begin{aligned} \frac{\delta \mathcal{F}}{\delta u} &= \frac{\partial f}{\partial u} - \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial u_x} \right) \\ &= u_x^2 - \frac{\partial}{\partial x} (2uu_x) \\ &= u_x^2 - (2u_x^2 + 2uu_{xx}) \\ &= -u_x^2 - 2uu_{xx}. \end{aligned}$$

If \mathcal{F} is a functional of more than one function (equivalently, of a vector function u), one may compute variational derivatives with respect to any of these (components).

$$\mathcal{F}[u, w] = \int_0^L u_x w dx, \quad \frac{\delta \mathcal{F}}{\delta w} = u_x, \quad \frac{\delta \mathcal{F}}{\delta u} = -\frac{\partial}{\partial x} \frac{\partial f}{\partial u_x} = -\frac{\partial}{\partial x} w = -w_x.$$

10. Hamiltonian partial differential equation. A Hamiltonian partial differential equation has the form

$$\frac{\partial u}{\partial t} = \mathcal{J} \frac{\delta \mathcal{H}}{\delta u},$$

where $u(t, x) : \mathcal{R} \times \mathcal{D} \rightarrow \mathcal{R}^d$. If $d = 1$, then we assume $u(t) \in \mathcal{V}$ for all $t \geq 0$, and if $d > 1$, we assume each component of u is an element of \mathcal{V} for all $t \geq 0$, i.e. $u(t) \in \mathcal{V}^d$, $t \geq 0$. The Hamiltonian is a functional $\mathcal{H}[u] : \mathcal{V}^d \rightarrow \mathcal{R}$. The operator \mathcal{J} is skew-symmetric with respect to the inner product $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{R}$. That is,

$$\langle u, \mathcal{J}v \rangle = -\langle v, \mathcal{J}u \rangle, \quad \forall u, v \in \mathcal{V}^d.$$

The operator \mathcal{J} may be a matrix (if $d > 1$) or a differential operator $\mathcal{J} = \frac{\partial}{\partial x}$ (check that this operator is skew symmetric for $d = 1$!) or a combination of these.

Let us consider an example. The sine-Gordon equation is

$$q_{tt} = q_{xx} - \sin q.$$

This equation can be written in the form of a Hamiltonian PDE by introducing the function $p = q_t$ and defining $u(t, x) = (q(t, x), p(t, x))^T$. The Hamiltonian structure is given by

$$\mathcal{H}[q, p] = \int_0^L \frac{p^2}{2} + \frac{q_x^2}{2} - \cos q \, dx, \quad \mathcal{J} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

for which

$$u_t = \mathcal{J} \frac{\delta \mathcal{H}}{\delta u} \iff \begin{pmatrix} q_t \\ p_t \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{pmatrix} \delta \mathcal{H} / \delta q \\ \delta \mathcal{H} / \delta p \end{pmatrix} \iff \begin{aligned} q_t &= p \\ p_t &= q_{xx} - \sin q. \end{aligned}$$

A second example is the simple advection equation $u_t = u_x$. This PDE has the rather obvious Hamiltonian structure:

$$\mathcal{H}[u] = \int_0^L \frac{u^2}{2} \, dx, \quad \mathcal{J} = \frac{\partial}{\partial x}.$$

The Hamiltonian is a conserved integral of a Hamiltonian PDE. To see this, compute

$$\frac{d}{dt} \mathcal{H}[u(t)] = \left\langle \frac{\delta \mathcal{H}}{\delta u}, u_t \right\rangle = \left\langle \frac{\delta \mathcal{H}}{\delta u}, \mathcal{J} \frac{\delta \mathcal{H}}{\delta u} \right\rangle = 0,$$

by the skew-symmetry of \mathcal{J} . For the previous example, we can check this:

$$\frac{d}{dt} \mathcal{H}[u(t)] = \frac{d}{dt} \int_0^L \frac{u^2}{2} \, dx = \int_0^L uu_t \, dx = \int_0^L uu_x \, dx = \int_0^L \frac{\partial}{\partial x} \left(\frac{u^2}{2} \right) \, dx = \frac{u^2}{2} \Big|_0^L = 0,$$

where the last equality follows from the periodic boundary conditions.

For the sine-Gordon equation we verify

$$\begin{aligned} \frac{d}{dt} \int_0^L \frac{p^2}{2} + \frac{q_x^2}{2} - \cos q \, dx &= \int_0^L pp_t + q_x q_{xt} + \sin q q_t \, dx, \\ &= \int_0^L p(q_{xx} - \sin q) + q_x p_x + p \sin q \, dx, \\ &= \int_0^L pq_{xx} + q_x p_x \, dx, \\ &= \int_0^L -p_x q_x + q_x p_x \, dx = 0, \end{aligned}$$

where we integrated by parts once.

Another conserved integral of the sine-Gordon equation is the linear momentum $\mathcal{M} = \int pq_x dx$. We verify:

$$\begin{aligned} \frac{d}{dt}\mathcal{M} &= \frac{d}{dt} \int_0^L pq_x dx, \\ &= \int_0^L p_t q_x + pq_{xt} dx, \\ &= \int_0^L (q_{xx} - \sin q)q_x + pp_x dx, \\ &= \int_0^L \frac{\partial}{\partial x} \left(\frac{q_x^2}{2} + \cos q + \frac{p^2}{2} \right) dx = 0, \end{aligned}$$

where, again, the periodic boundary conditions have been used to equate the last integral to zero.

One may verify that the Korteweg-de Vries equation (21) is a Hamiltonian PDE with

$$(23) \quad \mathcal{H}[u] = \int u^3 + \frac{u_x^2}{2} dx$$

and skew-symmetric form $\mathcal{J} = \frac{\partial}{\partial x}$.

11. Hamiltonian spatial discretization. The goal of discretization of Hamiltonian PDEs is to cast the discrete ODEs in the form of a Hamiltonian ODE, since this will imply that there is a conserved approximation of the Hamiltonian (total energy), and allows the use of symplectic time integrators. A secondary objective is the preservation of other conserved integrals. However, it may not be possible preserve all first integrals under discretization.

Our first step is to define the vector space upon which the solution will be defined. This is done by defining a mesh $x_i = i\Delta x$, $i = 0, \dots, N$, $\Delta x = L/N$. (In some applications it is useful to define a dual mesh as well, but that is beyond our scope here.) The discrete vector space $V \subset \mathcal{R}^N$ consists of functions on the mesh, denoted by $u(t) \in V$, where $u_i(t) \approx u(t, x_i)$, $i = 0, \dots, N-1$. Here the periodic boundary conditions mean we identify $u_N = u_0$, etc. We may also have vector valued grid functions $u \in V^d$.

Additionally we define an inner product on V^d :

$$\langle u, v \rangle = \sum_{i=0}^{N-1} u_i^T v_i \Delta x.$$

Given a function $F : V \rightarrow \mathcal{R}$, we can define the variational derivative with respect to the inner product $\langle \cdot, \cdot \rangle$,

$$(24) \quad \left\langle \frac{\delta F}{\delta u}, v \right\rangle = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (F(u + \varepsilon v) - F(u)), \quad \forall v \in V$$

We consider examples analogous to the first two examples in the continuous case. First,

$$F(u) = \sum_{i=0}^{N-1} u_i^3 \Delta x.$$

The variational derivative is

$$\begin{aligned} \left\langle \frac{\delta F}{\delta u}, v \right\rangle &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \sum_{i=0}^{N-1} [(u_i + \varepsilon v_i)^3 - u_i^3] \Delta x, \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \sum_{i=0}^{N-1} [u_i^3 + \varepsilon 3u_i^2 v_i + \varepsilon^2 3u_i v_i^2 + \varepsilon^3 v_i^3 - u_i^3] \Delta x, \\ &= \lim_{\varepsilon \rightarrow 0} \sum_{i=0}^{N-1} [3u_i^2 v_i + \varepsilon 3u_i v_i^2 + \varepsilon^2 v_i^3] \Delta x, \\ &= \sum_{i=0}^{N-1} 3u_i^2 v_i \Delta x. \end{aligned}$$

The last line is in the form of the inner product with v , so the variational derivative is $\delta F/\delta u = 3u^2$ (where the square is applied elementwise).

As a second example consider

$$F(u) = \sum_{i=0}^{N-1} \frac{1}{2} \left(\frac{u_{i+1} - u_i}{\Delta x} \right)^2 \Delta x.$$

The variational derivative is

$$\begin{aligned} \left\langle \frac{\delta F}{\delta u}, v \right\rangle &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \sum_{i=0}^{N-1} \left[\frac{1}{2} \left(\frac{u_{i+1} + \varepsilon v_{i+1} - u_i - \varepsilon v_i}{\Delta x} \right)^2 - \frac{1}{2} \left(\frac{u_{i+1} - u_i}{\Delta x} \right)^2 \right] \Delta x, \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \sum_{i=0}^{N-1} \frac{\Delta x}{2\Delta x^2} [(u_{i+1} - u_i)^2 + 2\varepsilon(u_{i+1} - u_i)(v_{i+1} - v_i) + \varepsilon^2(v_{i+1} - v_i)^2 - (u_{i+1} - u_i)^2], \\ &= \lim_{\varepsilon \rightarrow 0} \sum_{i=0}^{N-1} \frac{\Delta x}{2\Delta x^2} [2(u_{i+1} - u_i)(v_{i+1} - v_i) + \varepsilon(v_{i+1} - v_i)^2], \\ &= \sum_{i=0}^{N-1} \frac{1}{\Delta x^2} (u_{i+1} - u_i)(v_{i+1} - v_i) \Delta x. \end{aligned}$$

To cast this in the form of an inner product with arbitrary v , we use the periodic boundary conditions and note that

$$\sum_{i=0}^{N-1} v_{i+1} (u_{i+1} - u_i) \Delta x = \sum_{i=0}^{N-1} v_i (u_i - u_{i-1}) \Delta x.$$

Consequently,

$$\left\langle \frac{\delta F}{\delta u}, v \right\rangle = \sum_{i=0}^{N-1} \frac{1}{\Delta x^2} (u_i - u_{i-1} + u_i - u_{i+1}) v_i \Delta x = \sum_{i=0}^{N-1} \left(-\frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} \right) v_i \Delta x.$$

From this it follows that $\delta F/\delta u$ is given by the standard central approximation of the second derivative:

$$(\delta F/\delta u)_i = -\frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2}, \quad i = 0, \dots, N-1.$$

Construction of a Hamiltonian spatial discretization proceeds by choosing a quadrature $H(u)$ to approximate the functional $\mathcal{H}[u]$ (by a sum), and choosing an approximation J for the operator \mathcal{J} that is skew-symmetric with respect to the discrete inner product. Then the discretization is defined by

$$(25) \quad u_t = J \frac{\delta H}{\delta u},$$

or in terms of individual grid points,

$$\frac{d}{dt} u_i = \sum_{j=0}^{N-1} J_{ij} \frac{\delta H}{\delta u_j}.$$

As an example, consider the sine-Gordon equation. We introduce the quadrature

$$H = \sum_{i=0}^{N-1} \left[\frac{p_i^2}{2} + \frac{1}{2} \left(\frac{q_{i+1} - q_i}{\Delta x} \right)^2 - \cos q_i \right] \Delta x,$$

for which

$$\frac{\delta H}{\delta q_i} = -\frac{q_{i+1} - 2q_i + q_{i-1}}{\Delta x^2} + \sin q_i, \quad \frac{\delta H}{\delta p_i} = p_i.$$

The operator \mathcal{J} is approximated by

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

where I is the identity matrix of dimension N .

Define

$$u = \begin{pmatrix} q \\ p \end{pmatrix}.$$

Then the discretization is given by (25). In terms of components this becomes

$$\frac{d}{dt} q_i = p_i, \quad \frac{d}{dt} p_i = \frac{q_{i+1} - 2q_i + q_{i-1}}{\Delta x^2} - \sin q_i,$$

and the Hamiltonian H is a conserved quantity.

We would like to show that the discretization (25) indeed defines a Hamiltonian ODE (17) in general. We assume that the inner product $\langle \cdot, \cdot \rangle$ is related to the Euclidean metric by a symmetric positive definite matrix M :

$$\langle a, b \rangle = a^T M b, \quad \forall a, b \in V^d$$

From the definition of the discrete variational derivative (24), it follows that

$$\begin{aligned} \left\langle \frac{\delta F}{\delta u}, v \right\rangle &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (F(u + \varepsilon v) - F(u)), \\ \left(\frac{\delta F}{\delta u} \right)^T M v &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (F(u) + \varepsilon \nabla F(u) \cdot v + \mathcal{O}(\varepsilon^2) - F(u)) \\ &= \nabla F(u)^T v, \end{aligned}$$

which must hold for all $v \in V$. We conclude that

$$\frac{\delta F}{\delta u} = M^{-1} \nabla F(u).$$

Furthermore, the condition that J be skew-symmetric with respect to $\langle \cdot, \cdot \rangle$ implies

$$\langle a, Jb \rangle = -\langle Ja, b \rangle \quad \Rightarrow \quad a^T M J b = -a^T J^T M b, \quad \forall a, b \in V^d.$$

It follows that MJ (and consequently JM^{-1}) is a skew-symmetric matrix.

Hence we see that

$$u_t = J \frac{\delta H}{\delta u} = JM^{-1} \nabla H(u) = \tilde{J} \nabla H(u)$$

is Hamiltonian, with skew-symmetric structure matrix $\tilde{J} = JM^{-1}$.

To see that the Hamiltonian is conserved note that

$$\frac{dH}{dt} = \nabla H(u) \cdot \frac{du}{dt} = \left(\frac{\delta H}{\delta u} \right)^T M \frac{du}{dt} = \left\langle \frac{\delta H}{\delta u}, J \frac{\delta H}{\delta u} \right\rangle = 0$$

by the skew-symmetry of J with respect to $\langle \cdot, \cdot \rangle$.

In the example above our choice of inner product is just a scalar multiple of the usual vector inner product. For this reason, the variational derivative is just a scalar multiple of the gradient ∇H , and J is a skew-symmetric matrix. Clearly, the resulting discretization defines a Hamiltonian system, to which we can apply the time integration methods of the previous sections. The discrete Hamiltonian H will be conserved either exactly or in the sense of backward error analysis.

12. Full discretization of the KdV equation. We derive a Hamiltonian splitting method for the KdV equation. We propose the following quadrature for the Hamiltonian (23)

$$H = \sum_i \left[u_i^3 + \frac{1}{2} \left(\frac{u_{i+1} - u_i}{\Delta x} \right)^2 \right] \Delta x$$

The variational derivative of H is

$$\frac{\delta H}{\delta u_i} = 3u_i^2 - \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x}.$$

To discretize the operator ∂_x , we simply use central differences

$$J = \begin{bmatrix} 0 & \frac{1}{2\Delta x} & & -\frac{1}{2\Delta x} \\ -\frac{1}{2\Delta x} & \ddots & \ddots & \\ & \ddots & \ddots & \frac{1}{2\Delta x} \\ \frac{1}{2\Delta x} & & -\frac{1}{2\Delta x} & 0 \end{bmatrix}.$$

This construction yields the Hamiltonian ODE

$$\begin{aligned} \dot{u}_i &= \frac{1}{2\Delta x} \left[3u_{i+1}^2 - 3u_{i-1}^2 - \frac{u_{i+2} - 2u_{i+1} + u_i}{\Delta x^2} + \frac{u_i - 2u_{i-1} + u_{i-2}}{\Delta x^2} \right] \\ &= 3 \frac{u_{i+1}^2 - u_{i-1}^2}{2\Delta x} - \frac{\frac{1}{2}u_{i+2} - u_{i+1} + u_{i-1} + \frac{1}{2}u_{i-2}}{\Delta x^3} \end{aligned}$$

To construct an explicit (better said, “only linearly implicit”) splitting method, let us split the Hamiltonian into three terms $H = H_1 + H_2 + H_3$, where

$$H_1 = \sum_{i \text{ odd}} u_i^3 \Delta x, \quad H_2 = \sum_{i \text{ even}} u_i^3 \Delta x, \quad H_3 = \sum_i \frac{1}{2} \left(\frac{u_{i+1} - u_i}{\Delta x} \right)^2 \Delta x.$$

The variational derivatives of these Hamiltonians are, respectively

$$\frac{\delta H_1}{\delta u_i} = \begin{cases} 3u_i^2, & i \text{ odd} \\ 0, & i \text{ even}, \end{cases} \quad \frac{\delta H_2}{\delta u_i} = \begin{cases} 0, & i \text{ odd} \\ 3u_i^2, & i \text{ even}, \end{cases} \quad \frac{\delta H_3}{\delta u_i} = -\frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2}.$$

The equations of motion corresponding to the Hamiltonian H_1 are

$$\begin{aligned} \dot{u}_i &= \frac{3}{2\Delta x} (u_{i+1}^2 - u_{i-1}^2), & i \text{ even}, \\ \dot{u}_i &= 0, & i \text{ odd}. \end{aligned}$$

We define the time- Δt flow $u(t + \Delta t) = \Phi_{\Delta t}^1(u(t))$ corresponding to the exact solution of the above system

$$u_i(t + \Delta t) = \begin{cases} u_i(t) + \frac{3\Delta t}{2\Delta x} (u_{i+1}(t)^2 - u_{i-1}(t)^2), & i \text{ even}, \\ u_i(t), & i \text{ odd}. \end{cases}$$

Similarly, we define the odd flow (due to the even cubic terms in the Hamiltonian) to obtain $\Phi_{\Delta t}^2$.

The flow due to H_3 is linear and takes the form $\dot{u} = -D_3 u$ for discretization matrix D_3 associated with the finite difference formula

$$\dot{u}_i = -\frac{\frac{1}{2}u_{i+2} - u_{i+1} + u_{i-1} + \frac{1}{2}u_{i-2}}{\Delta x^3}.$$

(The matrix $D_3 = JD_2$, where D_2 is the standard second order central difference for the second derivative). We can either solve the above exactly, yielding the time- Δt map

$$\Phi_{\Delta t}^3 = \exp(-\Delta t D_3),$$

or we can use implicit midpoint rule

$$\tilde{\Phi}_{\Delta t}^3 = (I + \frac{\Delta t}{2} D_3)^{-1} (I - \frac{\Delta t}{2} D_3)$$

A first order time integrator is obtained by composing the flows

$$\Psi_{\Delta t} = \Phi_{\Delta t}^1 \circ \Phi_{\Delta t}^2 \circ \Phi_{\Delta t}^3.$$

A second order time integrator is obtained by symmetric composition

$$\Psi_{\Delta t} = \Phi_{\Delta t/2}^1 \circ \Phi_{\Delta t/2}^2 \circ \Phi_{\Delta t}^3 \circ \Phi_{\Delta t/2}^2 \circ \Phi_{\Delta t/2}^1.$$

The above method, with the implicit midpoint rule approximating $\Phi_{\Delta t}^3$, was used to compute the solution shown in [Figure 5](#).

13. Sources. These notes were compiled from material in the three excellent monographs. Symplectic Runge-Kutta methods are surveyed in:

- J.M. Sanz-Serna & M.P. Calvo, *Numerical Hamiltonian Problems*, Chapman-Hall, 1994.

More generally, methods that conserve first integrals and symmetries are referred to as “geometric integrators” and studied in

- E. Hairer, C. Lubich & G. Wanner, *Geometric Numerical Integration*, Springer-Verlag, Second Edition, 2006.
- B. Leimkuhler & S. Reich, *Simulating Hamiltonian Dynamics*, Cambridge University Press, 2005.

The latter book also contains a chapter on Hamiltonian PDEs. Specific ideas on spatial discretization of Hamiltonian PDEs are well explained in a number of papers by R. I. McLachlan, for instance (but there are many more):

- R. I. McLachlan, “Spatial discretization of PDEs with integrals”, *IMA J. Numer. Anal.* **23** (2003), 645–664.
- A. Kitson, R. I. McLachlan & N. Robidoux, “Skew-adjoint finite difference methods on nonuniform grids”, *New Zealand J. Math.*, **32** (2003) 139–159.