Advanced computational methods-Lecture 13

Symplectic methods and stochastic Hamiltonian system

1 Hamiltonian system and symplecticity

The Hamilton ODE is given by

$$\frac{dz}{dt} = J\nabla_z H$$

for $z = (p, q)^T$, where

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

Consider two (tangent) vectors in the phase space

$$\xi = \begin{pmatrix} \xi^p \\ \xi^q \end{pmatrix}, \ \eta = \begin{pmatrix} \eta^p \\ \eta^q \end{pmatrix}.$$

The parallelgram spanned by these two vectors are given by

$$P = \{t\xi + s\eta : 0 \le t \le 1, 0 \le s \le 1\}.$$

The projection of this parallelgram projected onto (p_i, q_i) plane is given by

$$\left\{ t \begin{pmatrix} \xi_i^p \\ \xi_i^q \end{pmatrix} + s \begin{pmatrix} \eta_i^p \\ \eta_i^q \end{pmatrix} : 0 \le t \le 1, 0 \le s \le 1 \right\},$$

whose area is given by $\xi_i^p \eta_i^q - \xi_i^q \eta_i^p$. Hence, the sum of all such areas is given by

$$\omega(\xi, \eta) = \sum_{i=1}^{d} \det \begin{pmatrix} \xi_i^p & \eta_i^p \\ \xi_i^q & \eta_i^q \end{pmatrix} = \sum_{i=1}^{d} (\xi_i^p \eta_i^q - \xi_i^q \eta_i^p) = \xi^T J \eta.$$

Definition 1. *A linear map A is called symplectic if*

$$A^T J A = J,$$

or $\omega(A\xi, A\eta) = \omega(\xi, \eta)$. A differentiable map is called symplectic if the Jacobian matrix is everywhere symplectic.

In other words, symplectic mapping preserves the sum of projected areas, or the quadratic forms.

A special class of symplectic mappings are **canonical transformation** $(p, q, t) \to (P, Q, t)$ such that the new variables (P, Q, t) form a new set of Hamilton equations (the Hamiltonian can be different).

1.1 Some descriptions of the symplectic mappings

We now consider a special transform

$$(p, q) \to (P, Q)$$

Note that $dP_i \wedge dQ_i$ is the area element for each componenet. The vectors, ξ, η are mapped into $A\xi, A\eta$ respectively. The old area is $\iint_D dp \wedge dq$ while the new area is $\iint dP \wedge dQ$. Hence, sum of area preserving means

$$dP \wedge dQ := \sum_i dP_i \wedge dQ_i = \sum_i dp_i \wedge dq_i =: dp \wedge dq.$$

The result below is a direct computation for which we omit the proof:

Proposition 1. *The condition $dP \wedge dQ = dp \wedge dq$ is equivalent to $\sum_i \frac{\partial(P_i, Q_i)}{\partial(p_\ell, p_k)} = 0$, $\sum_i \frac{\partial(P_i, Q_i)}{\partial(q_\ell, q_k)} = 0$, and $\sum_i \frac{\partial(P_i, Q_i)}{\partial(p_\ell, q_k)} = \delta_{k\ell}$, or*

$$P_p^T Q_p - Q_p^T P_p = 0, \quad P_p^T Q_q - Q_p^T P_q = I, \quad Q_q^T P_q - P_q^T Q_q = 0,$$

which is equivalent to

$$\Psi^T J \Psi = J,$$

where Ψ is the Jacobian matrix.

There is a useful lemma

Lemma 1. *Let D be a region.If f is continuously differentiable and ∇f is symmetric everywhere, then $f = \nabla H$ for some H.*

Assume $0 \in D$ (if not, consider $y - y_0$). One can define

$$H(y) = \int_0^1 y^T f(ty)\, dt.$$

One can then verify directly that $\nabla H = f$.

With the above observation, one has

Theorem 1. *The mapping $(p, q) \to (P, Q)$ is symplectic if and only if*

$$P^T dQ - p^T dq = dS$$

for some S.

Proof.

$$\begin{aligned} P^T dQ - p^T dq &= P^T Q_p dp + (P^T Q_q - p^T) dq \\ &= (Q_p^T P) \cdot dp + (Q_q^T P - p) \cdot dq \end{aligned}$$

This is a total differential is then equivalent to that the Jacobian matrix of the gradient is symmetric, which is equivalent to the conditions in Proposition 1. □

2 Symplectic numerical methods

2.1 The symplectic Euler method

Consider the following Euler scheme:

$$\begin{aligned} p_{n+1} &= p_n - \tau \nabla_q H(p_{n+1}, q_n), \\ q_{n+1} &= q_n + \tau \nabla_p H(p_{n+1}, q_n). \end{aligned}$$

If the Hamiltonian is separable, $H(p,q) = \frac{|p|^2}{2m} + U(q)$, this is reduced to the one we have seen in the HMC algorithm:

$$\begin{aligned} p_{n+1} &= p_n - \tau \nabla_q U(q_n), \\ q_{n+1} &= q_n + \tau \frac{p_{n+1}}{m}. \end{aligned}$$

The implicit scheme becomes explicit in this case!

We check that it is symplectic:

$$p_{n+1} \cdot dq_{n+1} = (p_n - \tau \nabla_q U(q_n)) \cdot (dq_n + \tau \frac{dp_{n+1}}{m})$$

Hence,

$$p_{n+1} \cdot dq_{n+1} - p_n \cdot dq_n = \frac{\tau}{m} p_n \cdot dp_{n+1} - \tau \nabla_q U(q_n) \cdot dq_n - \tau^2 \nabla_q U(q_n) \cdot \frac{dp_{n+1}}{m}$$

The second term is a total differential. For the first term, $p_n \cdot p_n$ is also a total differential, while $\nabla_q U \cdot d\nabla_q U$ is also total differential. Hence, the remaining term is

$$\frac{\tau}{m} p_n \cdot (-\tau d\nabla_q U(q_n)) - \frac{\tau^2}{m} \nabla_q U(q_n) \cdot dp_n,$$

which again makes the total differential.

From here, we find that the verification in this way is not very simple. In fact, one can also chooses to check $Q \cdot dP + p \cdot dq$ and see whether it is total differential...

A more direct way is to to check $dp_{n+1} \wedge dq_{n+1}$ or compute the determinants in Proposition 1.

Exercise: check these conditions.

2.2 Symplectic Runge-Kutta methods

Consider the ODE

$$\dot{y} = f(y).$$

Now, consiser the Runge-Kutta methods

$$y_{n+1} = y_n + \tau \sum_{i=1}^{s} b_i K_i,$$

$$K_i = f(t_n + c_i \tau, y_n + h \sum_{j=1}^{s} a_{ij} K_j)$$

Proposition 2. *If for any f such that the ODE has a quadratic first integral $I(y) = y^T C y$ for some symmetric C the Runga-Kutta method converve this first integral, then this method is symplectic when applied on Hamiltonian flows.*

Note that the assumption on the symmetry of C is not important, because any quadratic form can be written as $y^T C y$ for some symmetric C.

The proof divides into two steps.

Step 1: When we apply the method to $\dot{y} = f(y)$ with $y(0) = y_0$ and obtain y_n, we can compute

$$\bar{\Psi}_n = \frac{\partial y_n}{\partial y_0}$$

We can also apply the method directly to $\dot{y} = f(y)$ and $\dot{\Psi} = f'(y)\Psi$. One can claim that

$$\Psi_n = \bar{\Psi}_n.$$

In other words, the partial derivative on y_0 and the RK method commute.

Step 2: If the ODE flow itself is symplectic, or

$$I(\Psi) = \Psi^T J \Psi$$

is a constant matrix, then Runga-Kutta method can preserve this quantity with property in the statement. Note that here J is not symmetric, how can we apply the conditions in the statement? Note that here we have matrices Ψ instead of vectors. For vectors, if J is antisymmetric, the quadratic form is zero. What we do is to check each compoenents of $\Psi^T J \Psi$, which are quadratic forms of the components of Ψ. Formulating (y, Ψ) into a big vector, each component can be written as $z^T C z$ for some symmetric C. The numerical method conserves quadratic form, and thus each component of $\Psi^T J \Psi$.

We omit the details of the proof.

With the above observation, one can derive the following sufficient condition of symplecticity of RK.

Theorem 2. *If the coefficients of the RK method satisfies*

$$b_i a_{ij} + b_j a_{ji} = b_i b_j, \quad \forall i, j$$

then the method is symplectic.

Proof. It sufficies to show that if there exists C symmetric such that

$$y^T C f(y) = 0, \ \forall y,$$

then $I_n := y_n^T C y_n$ is a constant.

This then can be shown directly. Using $y_{n+1} = y_n + \tau \sum_{i=1}^s b_i K_i$ and the symmetry of C, we derive directly that

$$y_{n+1}^T C y_{n+1} = y_n^T C y_n + \tau 2 \sum_{i=1}^s b_i y_n^T C K_i + \tau^2 \sum_{i,j} b_i b_j K_i^T C K_j$$

Next, one notice $K_i = f(Y_i)$ and thus

$$y_n^T C f(Y_i) = (Y_i - \tau \sum_{j=1}^s a_{ij} K_j)^T C f(Y_i) = -\tau \sum_{j=1}^s a_{ij} K_j^T C K_i.$$

Inserting this back, one finds that the claim holds. □

2.3 Gauss collocation method

Will not go over in class.

The collocation method is to choose several points and the impose the conditions to hold on these points **exactly**. This is like the spirit of interpolation. In fact, the interpolation can be viewed as a type of collocation method.

For ODEs, the collocation method is as follows: pick a polynomial $u(t)$ such that $u(t_n) = y_n$. Then, impose the conditions

$$\dot{u}(t_n + c_i \tau) = f(t_n + c_i \tau, u(t_n + c_i \tau)), \quad i = 1, \cdots, s,$$

so that u can be determined. Then, one lets

$$y_{n+1} = u(t_n + \tau).$$

The Gaussian collocation method is as follows: let $c_1, \cdots, c_s$ be the s zeros of $\frac{d^s}{dx^s}(x^s(1-x)^s)$, and u be a polynomial of degree s. Then, the method is determined. This is called the Gauss collocation method.

Theorem 3. *The Gaussian collocation method conserves all quadratic first integrals and hence symplectic.*

Proof. Let $I(y) = y^T C y$ be a first integral (for any path) so that

$$y^T C f(y) = 0, \quad \forall y.$$

Then, for the collocation method,

$$y_{n+1}^T C y_{n+1} - y_n^T C y_n = 2 \int_{t_n}^{t_{n+1}} u^T(t) C \dot{u} \, dt.$$

However, for the polynomial with degree $2s-1$, the Gaussian quadrature is exact. Hence,

$$\int_{t_n}^{t_{n+1}} u^T(t) C \dot{u} \, dt = \sum_{i=1}^{s} w_i u^T(t_n + c_i \tau) C \dot{u}(t_n + c_i \tau)$$

Using the condition,

$$u^T(t_n + c_i\tau) C \dot{u}(t_n + c_i\tau) = u^T(t_n + c_i\tau) C f(t_n + c_i\tau, u(t_n + c_i\tau)) = 0.$$

This means that the integral is zero. Hence, the quadratic quantity is conserved. □

3 Stochastic Hamiltonian system and a symplectic scheme

Consider the stochastic system

$$dP = f(P,Q)\, dt + \sum_{r=1}^{m} \sigma_r(P,Q) \circ dW_r,$$
$$dQ = g(P,Q)\, dt + \sum_{r=1}^{m} \gamma_r(P,Q) \circ dW_r.$$

Here, "$\circ$" means the stochastic integral is in Stratonovich sense. Here, the W_r process in the two SDEs are the **same**.

We consider the Hamiltonian system

$$f = -\nabla_q H_0, \quad g = \nabla_p H_0,$$
$$\sigma_r = -\nabla_q H_r, \quad \gamma_r = \nabla_p H_r, \quad r = 1, \cdots, m.$$

Proposition 3. *The mapping $(p,q) \to (P(t), Q(t))$ is differentiable almost surely (of course, not differentiable in t), and is symplectic almost surely.*

The proof of differentiability is beyond this course. If you are interested, you may read corresponding books.

Here, we only check the symplecticity. We have seen that the symplecticity is equivalent to the following conditions:

$$\sum_i \frac{\partial(P_i, Q_i)}{\partial(p_\ell, p_k)} = 0, \quad \sum_i \frac{\partial(P_i, Q_i)}{\partial(q_\ell, q_k)} = 0, \quad \sum_i \frac{\partial(P_i, Q_i)}{\partial(p_\ell, q_k)} = \delta_{k\ell}.$$

To check these conditions, we use Itô's formulas to check that their differentials are zero. For example, for the first condition, we need to check

$$\sum_i d\left(\frac{\partial P_i}{\partial p_\ell}\frac{\partial Q_i}{\partial p_k} - \frac{\partial P_i}{\partial p_k}\frac{\partial Q_i}{\partial p_\ell}\right) = 0.$$

The other two are similar. Below, we only focus on the first as an example.

3.1 Stratonovich Integral and the Itô formula

Consider

$$dX = b(X)\,dt + \sigma(X) \circ dW.$$

Consider the Stratonovich SDEs in $\mathbb{R}^d$

$$dX = b(X)dt + \sigma(X) \circ dW$$

If σ is a smooth function, this can be converted into SDEs in Itô sense as

$$dX = (b(X) + \bar{b}(X))dt + \sigma(X)dW,$$

with

$$\bar{b}_j(x) = \frac{1}{2}\partial_m \sigma_{jk}(x)\sigma_{mk}(x).$$

To understand this formally, we consider the Riemann sum for approximating the stochastic integral in Stratonovich sense

$$\int_0^T \sigma(X) \circ dW \approx \sum_i \sigma(X_{t_{i+1/2}})(W_{t_{i+1}} - W_{t_i}).$$

Approximately, we have

$$\begin{aligned}\sigma_{jk}(X_{t_{i+1/2}})(W^k_{t_{i+1}} - W^k_{t_i}) &= \sigma_{jk}(X_{t_i})(W^k_{t_{i+1}} - W^k_{t_i}) \\ &+ \partial_m \sigma_{jk}(X_{t_i})\sigma_{mn}(W^n_{t_{i+1/2}} - W^n_{t_i})(W^k_{t_{i+1}} - W^k_{t_i}) + \text{higher order terms}\end{aligned}$$

Since the Brownian motion has independent increments and has quadratic variation as t, $(W^n_{t_{i+1/2}} - W^n_{t_i})(W^k_{t_{i+1}} - W^k_{t_i})$ can be approximated by $\frac{1}{2}\Delta t_i \delta_{nk}$.

The Itô's formula for X will also be changed:

$$df(X) = dX \cdot \nabla f(X) + \frac{1}{2}\sigma\sigma^T : \nabla^2 f\, dt = dX \circ \nabla f(X).$$

Consequently, we have

$$d(XY) = X \circ dY + Y \circ dX.$$

3.2 Verification of simplecticity

For convenience, we introduce

$$P^{ik}_p = \frac{\partial P^i}{\partial p^k}, P^{ik}_q = \frac{\partial P^i}{\partial q^k}.$$

Similarly, one can define Q^{ik}_p and Q^{ik}_q.

We need to verify

$$d(P^{i\ell}_p Q^{ik}_p - P^{ik}_p Q^{i\ell}_p) = 0.$$

Using the equation for P, one finds

$$dP^{ik}_p = \sum_{\alpha=1}^{d}(\frac{\partial f^i}{\partial P^\alpha} P^{\alpha k}_p + \frac{\partial f^i}{\partial Q^\alpha} Q^{\alpha k}_p)\, dt + \sum_{r=1}^{m}\sum_{\alpha=1}^{d}(\frac{\partial \sigma^i_r}{\partial P^\alpha} P^{\alpha k}_p + \frac{\partial \sigma^i_r}{\partial Q^\alpha} Q^{\alpha k}_p) \circ dW_r.$$

Similarly, you can find equation for other variables. Then, apply the Itô's formula above, one can find that the conditions become

$$\sum_{i=1}^{d}\sum_{j=1}^{d}(\frac{\partial f^i}{\partial P^j} P^{jk}_p Q^{i\ell}_p + \frac{\partial f^i}{\partial Q^j} Q^{jk}_p Q^{i\ell}_p + \frac{\partial g^i}{\partial P^j} P^{j\ell}_p P^{ik}_p + \frac{\partial g^i}{\partial Q^j} Q^{j\ell}_p P^{ik}_p$$
$$-\frac{\partial f^i}{\partial P^j} P^{j\ell}_p Q^{ik}_p - \frac{\partial f^i}{\partial Q^j} Q^{j\ell}_p Q^{ik}_p - \frac{\partial g^i}{\partial P^j} P^{jk}_p P^{i\ell}_p - \frac{\partial g^i}{\partial Q^j} Q^{jkl}_p P^{i\ell}_p) = 0.$$

and

$$\sum_{i=1}^{d}\sum_{j=1}^{d}(\frac{\partial \sigma^i_r}{\partial P^j} P^{jk}_p Q^{i\ell}_p + \frac{\partial \sigma^i_r}{\partial Q^j} Q^{jk}_p Q^{i\ell}_p + \frac{\partial \gamma^i_r}{\partial P^j} P^{j\ell}_p P^{ik}_p + \frac{\partial \gamma^i_r}{\partial Q^j} Q^{j\ell}_p P^{ik}_p$$
$$-\frac{\partial \sigma^i_r}{\partial P^j} P^{j\ell}_p Q^{ik}_p - \frac{\partial \sigma^i_r}{\partial Q^j} Q^{j\ell}_p Q^{ik}_p - \frac{\partial \gamma^i_r}{\partial P^j} P^{jk}_p P^{i\ell}_p - \frac{\partial \gamma^i_r}{\partial Q^j} Q^{jkl}_p P^{i\ell}_p) = 0.$$

These conditions clearly hold for the stochastic Hamiltonian systems.

3.3 A numerical scheme and the symplecticity

We first change the stochastic Hamiltonian system into the Itô's form:

$$
\begin{aligned}
dP &= f(P,Q)\,dt + \frac{1}{2}\sum_{r=1}^{m}(\sigma_r \cdot \nabla_P \sigma_r + \gamma_r \cdot \nabla_Q \sigma_r)\,dt + \sum_{r=1}^{m}\sigma_r(P,Q)dW_r,\\
dQ &= g(P,Q)\,dt + \frac{1}{2}\sum_{r=1}^{m}(\sigma_r \cdot \nabla_P \gamma_r + \gamma_r \cdot \nabla_Q \gamma_r)\,dt + \sum_{r=1}^{m}\gamma_r(P,Q)dW_r.
\end{aligned}
$$

Now, we apply the scheme where P is implicit while Q is explicit. Hoping that this will be symplectic. The method then gives

$$
\begin{aligned}
P_{n+1} =& P_n + f(P,Q)\,h + \frac{1}{2}\sum_{r=1}^{m}(\sigma_r \cdot \nabla_P \sigma_r + \gamma_r \cdot \nabla_Q \sigma_r)\,h\\
&- \sum_{r=1}^{m}(\sigma_r \cdot \nabla_P \sigma_r)h + \sum_{r=1}^{m}\sigma_r(P_{n+1},Q_n)\sqrt{h}\zeta_{n,r},\\
Q_{n+1} =& Q_n + g(P,Q)\,h + \frac{1}{2}\sum_{r=1}^{m}(\sigma_r \cdot \nabla_P \gamma_r + \gamma_r \cdot \nabla_Q \gamma_r)\,h\\
&- \sum_{r=1}^{m}(\sigma_r \cdot \nabla_P \gamma_r)h + \sum_{r=1}^{m}\gamma_r(P_{n+1},Q_n)\sqrt{h}\zeta_{n,r},
\end{aligned}
$$

where we recall the truncated variable

$$
\zeta_r = \begin{cases} \xi_r, & |\xi_r| \le A_h,\\ sgn(\xi_r)A_h, & else \end{cases}
$$

We have the following claim

Theorem 4. *The method is of strong order* $1/2$ *and is symplectic, for the stochastic Hamiltonian system.*

Proof. The scheme in the stochastic Hamiltonian case is given by

$$\begin{aligned}
P_{n+1} =& P_n - \nabla_Q H_0(P_{n+1}, Q_n)h - \frac{1}{2}\sum_{r=1}^{m} \nabla_Q \nabla_P H_r \cdot \nabla_Q H_r\, h \\
&- \frac{1}{2}\sum_{r=1}^{m} \nabla_Q^2 H_r \cdot \nabla_P H_r\, h - \sum_{r=1}^{m} \nabla_Q H_r \zeta_r \sqrt{h}, \\
Q_{n+1} =& Q_n + \nabla_P H_0(P_{n+1}, Q_n)h + \frac{1}{2}\sum_{r=1}^{m} \nabla_P \nabla_P H_r \cdot \nabla_Q H_r\, h \\
&+ \frac{1}{2}\sum_{r=1}^{m} \nabla_P \nabla_Q H_r \cdot \nabla_P H_r\, h + \sum_{r=1}^{m} \nabla_P H_r \zeta_r \sqrt{h}.
\end{aligned}$$

This scheme can be written as

$$\begin{aligned}
P_{n+1} &= P_n - \nabla_Q F(P_{n+1}, Q_n), \\
Q_{n+1} &= Q_n + \nabla_P F(P_{n+1}, Q_n)
\end{aligned}$$

where

$$F(P,Q) = H_0(P,Q)h + \frac{1}{2}\sum_{r=1}^{m} \nabla_Q H_r \cdot \nabla_P H_r h + \sum_{r=1}^{m} H_r \zeta_r \sqrt{h}$$

One can see that this is already in the Hamiltonian flow. We know this scheme is for deterministic flow and thus we conclude that this is symplectic for the stochastic Hamiltonian system. (See below.) □

Lemma 2. *Let F be a C^1 function.*

$$\begin{aligned}
P_{n+1} &= P_n - \nabla_Q F(P_{n+1}, Q_n), \\
Q_{n+1} &= Q_n + \nabla_P F(P_{n+1}, Q_n).
\end{aligned}$$

Then,

$$dP_{n+1} \wedge dQ_{n+1} = dP_n \wedge dQ_n.$$

Proof. By direct computation,

$$\begin{aligned}
dP_{n+1} \wedge dQ_{n+1} = \sum_{i=1}^{d} dP_{n+1}^i \wedge dQ_{n+1}^i &= \sum_i dP_{n+1}^i \wedge (dQ_n^i + \sum_j \partial_{P_i P_j} F dP_{n+1}^j + \sum_j \partial_{Q_j P_i} F dQ_n^j) \\
&= \sum_i dP_{n+1}^i \wedge (dQ_n^i + \sum_j \partial_{Q_j P_i} F dQ_n^j)
\end{aligned}$$

Moreover,

$$dP^i_{n+1} = dP^i_n - \sum_{j=1}^{d} \partial_{P_j}\partial_{Q_i} F \, dP^j_{n+1} - \sum_{j=1}^{d} \partial_{Q_j}\partial_{Q_i} F \, dQ^j_n$$

Hence,

$$\begin{aligned}\sum_i dP^i_{n+1} \wedge dQ^i_n &= \sum_i (dP^i_n - \sum_{j=1}^{d} \partial_{P_j}\partial_{Q_i} F \, dP^j_{n+1} - \sum_{j=1}^{d} \partial_{Q_j}\partial_{Q_i} F \, dQ^j_n) \wedge dQ^i_n \\ &= \sum_i dP^i_n \wedge dQ^i_n - \sum_{j=1}^{d} \partial_{P_j}\partial_{Q_i} F \, dP^j_{n+1} \wedge dQ^i_n\end{aligned}$$

The second term here happens to cancel the second term above. Hence, the claim is proved. □