Symplectic Runge-Kutta methods for nonsymplectic problems

J. M. Sanz-Serna
Universidad Carlos III de Madrid

I. A LITTLE BACKGROUND

RUNGE-KUTTA (RK) METHODS 1895, 1901

• Specified by s (number of stages) and s^2+2s real numbers a_{ij}, $i,j=1,\ldots,s$, b_i, c_i, $i=1,\ldots,s$.

• Integrates D-dimensional system $(d/dt)y = F(y,t)$ by time-stepping $n \to n+1$, $n=0,\ldots,N-1$, ($h_n = t_{n+1}-t_n$ is steplength)

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

where $K_{n,i} = F(Y_{n,i}, t_n + c_i h_n)$ are the 'slopes' at the so-called internal stages $Y_{n,i}$. These are defined by

$$Y_{n,i} = y_n + h_n \sum_{j=1}^{s} a_{ij} K_{n,j}, \quad i=1,\ldots,s.$$

(Recursion/algebraic system.)

QUADRATIC INVARIANTS (eg angular momentum)

- Some RK exactly conserve quadratic invariants:

Theorem 1. *(Cooper 1987) If the system possesses a quadratic first integral I [ie $I(\cdot,\cdot)$ is a bilinear map $\mathbb{R}^D \times R^D \to \mathbb{R}$ and along each solution $(d/dt)I(y(t),y(t)) \equiv 0$] and*

$$b_i a_{ij} + b_j a_{ji} - b_i b_j = 0, \qquad i,j = 1,\ldots,s,$$

then for each RK trajectory $\{y_n\}$, $I(y_n, y_n)$ is independent of n.

(The condition linking the a_{ij}, b_i is essentially necessary for conclusion to hold.)

(The proof is a simple algebraic manipulation.)

HAMILTONIAN PROBLEMS

• Appear very frequently. They are characterized by the fact that their solution flow preserves the (canonical) symplectic structure of the phase space.

• A numerical integrator is said to be *symplectic* if when applied to Hamiltonian systems the map $y_n \mapsto y_{n+1}$ preserves the symplectic structure of the phase space.

• Symplectic integrators to be preferred in many applications. Extensive theory developed in last 20 years.

SYMPLECTIC RK METHODS

Theorem 2. *(Lasagni, Suris, SS 1988) If*

$$b_i a_{ij} + b_j a_{ji} - b_i b_j = 0, \qquad i, j = 1, \ldots, s,$$

then the RK method is symplectic.

(The condition linking the a_{ij}, b_i is essentially necessary for conclusion to hold. Methods that satisfy such a condition are called symplectic RK methods.)

(The proof is a simple algebraic manipulation.)

PARTITIONED RUNGE-KUTTA (PRK) METHODS

• Some systems come to us *partitioned* (eg positions/velocities) $(d/dt)q = f(q,p,t)$, $(d/dt)p = g(q,p,t)$.

• Makes sense to use two sets of RK coefficients a_{ij}, $i,j = 1,\ldots,s$, b_i, c_i,$i = 1,\ldots,s$ AND A_{ij}, $i,j = 1,\ldots,s$, B_i, C_i,$i = 1,\ldots,s$. Result is a Partitioned Runge-Kutta scheme (PRK).

• The Verlet/Stormer/leapfrog method (method of choice in molecular dynamics) gives an example of PRK method.

PARTITIONED RUNGE-KUTTA (PRK) METHODS (continued)

- Equations are now:

$$q_{n+1} = q_n + h_n \sum_{i=1}^{s} b_i k_{n,i}, \qquad p_{n+1} = p_n + h_n \sum_{i=1}^{s} B_i \ell_{n,i},$$

with slopes

$$k_{n,i} = f(Q_{n,i}, P_{n,i}, t_n + c_i h_n), \; \ell_{n,i} = g(Q_{n,i}, P_{n,i}, t_n + C_i h_n),$$

and stages

$$Q_{n,i} = q_n + h_n \sum_{i=1}^{s} a_{ij} k_{n,j}, \qquad P_{n,i} = p_n + h_n \sum_{j=1}^{s} A_{ij} \ell_{n,j}.$$

CONSERVATION PROPERTIES OF PRK METHODS

- Both theorems above easily adapted:

Theorem 3. *$S(\cdot,\cdot)$ is a real-valued bilinear map such that along solutions*

$$(d/dt)S(q(t),p(t)) \equiv 0.$$

The relations

$$b_i = B_i,\ i = 1,\ldots,s, \quad b_iA_{ij} + B_ja_{ji} - b_iB_j = 0,\ i,j = 1,\ldots,s,$$

and

$$c_i = C_i, \quad i = 1,\ldots,s,$$

guarantee that, for each PRK trajectory $S(q_n,p_n)$ is independent of n. The same relations ensure that the PRK method is symplectic (Abia & SS 1989, 1993, Suris 1990).

II. VARIATIONAL EQUATIONS/ADJOINTS/ SENSITIVITIES

VARIATIONAL SYSTEM

- d-dimensional system $(d/dt) = f(x,t)$, initial cndtn. $x(t_0) = \alpha$.

- Let $\bar{x}(t)$ be solution with $\bar{x}(t_0) = \alpha + \eta$, η 'small'. Wish to estimate $\bar{x}(t_0 + T) - x(t_0 + T)$.

- Note:

$$(d/dt)(\bar{x}(t) - x(t)) = f(\bar{x}(t), t) - f(x(t), t) \approx \partial_x f(x(t), t)(\bar{x}(t) - x(t)).$$

- As $|\eta| \to 0$,

$$\bar{x}(t) = x(t) + \delta(t) + o(|\eta|),$$

where δ solves the variational equations:

$$\frac{d}{dt}\delta = \partial_x f(x(t), t)\, \delta.$$

THE ADJOINT SYSTEM

- The adjoint system

$$\frac{d}{dt}\lambda = -\partial_x f(x(t),t)^{\mathsf{T}}\,\lambda$$

is set up so as to get

$$\lambda(t_0+T)^{\mathsf{T}}\delta(t_0+T) = \lambda(t_0)^{\mathsf{T}}\delta(t_0).$$

(Note

$$\frac{d}{dt}\lambda(t)^{\mathsf{T}}\delta(t) = \left(\frac{d}{dt}\lambda(t)\right)^{\mathsf{T}}\delta(t) + \lambda(t)^{\mathsf{T}}\left(\frac{d}{dt}\delta(t)\right) \equiv 0.)$$

USING THE ADJOINT SYSTEM

• Assume we wish to compute $\omega^{\mathsf{T}}\delta(t_0+T)$. Solve adjoint with final condition $\lambda(t_0+T)=\omega$ to find $\lambda(t_0)$ and then

$$\omega^{\mathsf{T}}\delta(t_0+T)=\lambda(t_0+T)^{\mathsf{T}}\delta(t_0+T)=\lambda(t_0)^{\mathsf{T}}\delta(t_0)=\lambda(t_0)^{\mathsf{T}}\eta.$$

• If η varies only *one* integration of adjoint! (Variational equations require a fresh integration for each new initial condition.)

• Eg $\nabla_\alpha \mathcal{C}(x(t_0+T))$ (real-valued $\mathcal{C}$). Use $\omega=\nabla_x\mathcal{C}(x(t_0+T))$ and η the unit vectors to conclude $\nabla_\alpha\mathcal{C}(x(t_0+T))=\lambda(t_0)$. Adjoint pulls back gradients.

• Variational eqns propagate perturbations forward, adjoint eqn work backward.

FINDING THE SENSITIVITY OF THE RK SOLUTION

When $x(t_0 + T)$ is found approximately by RK method, do we find the variation of the computed $x_N \approx x(t_0 + T)$?

* By numerically integrating the variational equation?

** By taking variations in the (discrete) RK equations?

FINDING THE SENSITIVITY OF THE RK SOLUTION (continued)

- Both approaches lead to the same thing.

- For Euler's rule (1768):

$$x_{n+1} = x_n + h_n f(x_n, t_n).$$

If we vary the discrete equations, we get:

$$\delta_{n+1} = \delta_n + h_n \partial_x f(x_n, t_n) \cdot \delta_n.$$

. . . and these two formulae coincide with Euler's rule as applied to the $2d$-dimensional system (original+variational) satisfied by (x, δ).

SENSITIVITY OF THE RK SOLUTION: ADJOINTS

- Recall usefulness of adjoint system hinges on

$$\omega^{\mathsf{T}}\delta(t_0+T) = \lambda(t_0+T)^{\mathsf{T}}\delta(t_0+T) = \lambda(t_0)^{\mathsf{T}}\delta(t_0) = \lambda(t_0)^{\mathsf{T}}\eta.$$

To compute the discrete sensitivity $\omega^{\mathsf{T}}\delta_N$ in terms of the dual variable λ, we likewise need λ_N such that $\lambda_N^{\mathsf{T}}\delta_N = \lambda_0^{\mathsf{T}}\delta_0$.

- This target will be achieved if $3d$-system for x, δ, λ is integrated by symplectic RK scheme, because this will exactly preserve the quadratic invariant $\lambda(t)^{\mathsf{T}}\delta(t)$ of the differential system.

- If integrator is not symplectic, integrating the adjoint equation results in λ_N such that $\lambda_N^{\mathsf{T}}\delta_N = \lambda_0^{\mathsf{T}}\delta_0 + \mathcal{O}(h^\nu)$ (ν order of method).

- Of course equations for δ are left out when finding λ_N.

THE PRK CASE

• $\lambda_N^{\mathsf{T}}\delta_N = \lambda_0^{\mathsf{T}}\delta_0$ also guaranteed if symplectic PRK is used (use lower case coefficients for x, δ and upper case coefficients for λ.

• If x system is integrated by nonsymplectic RK method with nonzero b_i's, define clever upper case coefficients

$$A_{ji} = b_i - b_i a_{ij}/b_j,\ i,j = 1,\ldots,s, \quad B_i = b_i, C_i = c_i,\ i = 1,\ldots,s,$$

and use them to integrate adjoint system. Overall integrator is symplectic and thus $\lambda_N^{\mathsf{T}}\delta_N = \lambda_0^{\mathsf{T}}\delta_0$.

• Eg, for nonsymplectic Euler's rule, the clever formula is

$$\lambda_{n+1} = \lambda_n - h_n \partial_x f(x_n, t_n)^{\mathsf{T}} \lambda_{n+1}.$$

(x_n found first from $x_0 = \alpha$ by stepping $n \to n+1$. Then get λ_n from $\lambda_N = \omega$ by stepping $n \leftarrow n+1$. Whole business is explicit.)

AN OBVIOUS QUESTION

• As we have just seen, we may compute exactly $\nabla_\alpha \mathcal{C}(x_N)$ by a clever integration of the adjoint system.

• Can't we find that gradient directly (ie by differentiation of the RK formulae that we used to find x_N starting from α)?

• Before we do that we need to take a small detour. . .

III. LAGRANGE MULTIPLIERS/ AUTOMATIC DIFFERENTIATION

GRADIENTS VIA LAGRANGE MULTIPLIERS: ODEs

New derivation of $\nabla_\alpha \mathcal{C}(x(t_0+T)) = \lambda(t_0)$:

- Set Lagrangian $\mathcal{L}$ with multipliers $\hat{\lambda}_0$, $\hat{\lambda}$ (for solns $\mathcal{L} = \mathcal{C}$):

$$\mathcal{C}(\hat{x}(t_0+T)) - \hat{\lambda}_0^{\mathsf{T}}\Big(\hat{x}(t_0) - \hat{\alpha}\Big) - \int_{t_0}^{t_0+T} \hat{\lambda}(t)^{\mathsf{T}}\left(\frac{d}{dt}\hat{x}(t) - f(\hat{x}(t),t)\right) dt,$$

- Take variations and integrate by parts:

$$\begin{aligned}
\delta\mathcal{L} \;=\; & \Big(\nabla_x \mathcal{C}(\hat{x}(t_0+T)) - \hat{\lambda}(t_0+T)\Big)^{\mathsf{T}} \delta(t_0+T) + \hat{\lambda}(t_0)^{\mathsf{T}}\eta \\
& + \Big(\hat{\lambda}(t_0) - \hat{\lambda}_0\Big)^{\mathsf{T}} \delta(t_0) \\
& + \int_{t_0}^{t_0+T} \left(\frac{d}{dt}\hat{\lambda}(t)^T) + \hat{\lambda}(t)^T \partial_x f(\hat{x}(t),t)\right)\delta(t)\, dt.
\end{aligned}$$

- If red parts vanish, $\delta\mathcal{L} = \lambda(t_0)^{\mathsf{T}}\eta$ and $\lambda(t_0)$ is the gradient sought.

A MORE COMPLICATED LAGRANGIAN

- May also use the slope k as an additional independent function:

$$\begin{aligned}\mathcal{L} \;=\; & \mathcal{C}(x(t_0+T)) - \lambda_0^{\mathsf{T}}\Big(x(t_0)-\alpha\Big) \\ & - \int_{t_0}^{t_0+T} \lambda(t)^{\mathsf{T}}\left(\frac{d}{dt}x(t) - k(t)\right) dt \\ & - \int_{t_0}^{t_0+T} \Lambda(t)^{\mathsf{T}}\Big(k(t) - f(x(t),t)\Big)\, dt.\end{aligned}$$

GRADIENTS VIA LAGRANGE MULTIPLIERS: MAPPINGS

- Consider $\psi(\alpha) = \Psi(\alpha, \gamma(\alpha))$, for some $\Psi : \mathbb{R}^{d+d'} \to \mathbb{R}$ with $\gamma(\alpha)$ defined by $\Omega(\alpha, \gamma) = 0$.

- In analogy with the preceding slide, introduce Lagrangian

$$\mathcal{L}(\alpha, \gamma, \lambda) = \Psi(\alpha, \gamma) + \lambda^T \Omega(\alpha, \gamma).$$

- Then:

$$\nabla_\alpha \psi|_{\alpha_0} = \nabla_\alpha \mathcal{L}(\alpha, \gamma, \lambda)|_{(\alpha_0, \gamma_0, \lambda_0)}.$$

with γ_0 and λ_0 defined by

$$\nabla_\lambda \mathcal{L}(\alpha, \gamma, \lambda)|_{(\alpha_0, \gamma_0, \lambda_0)} = 0, \quad \nabla_\gamma \mathcal{L}(\alpha, \gamma, \lambda)|_{(\alpha_0, \gamma_0, \lambda_0)} = 0.$$

AN EXAMPLE: $\nabla\psi(\alpha_1,\alpha_2) = \cos^2\alpha_1 + \cos\alpha_1\cos\alpha_2 + \sin^3\alpha_2$

Intermediate variables γ_j defined by equations
$\Omega_1 = \gamma_1 - \cos\alpha_1 = 0$, $\Omega_2 = \gamma_2 - \cos\alpha_2 = 0$, $\Omega_3 = \gamma_3 - \sin\alpha_2 = 0$.

Lagrangian:

$$\begin{aligned}\mathcal{L} \;=\; & \gamma_1^2 + \gamma_1\gamma_2 + \gamma_3^3 \\ & + \lambda_1(\gamma_1 - \cos\alpha_1) + \lambda_2(\gamma_2 - \cos\alpha_2) + \lambda_3(\gamma_3 - \sin\alpha_2).\end{aligned}$$

Differentiate wrt λ_i and get equations $\Omega_i = 0$.
Differentiate wrt γ_i: $2\gamma_1 + \gamma_2 + \lambda_1 = 0$, $\gamma_1 + \lambda_2 = 0$, $\gamma_3^2 + \lambda_3 = 0$.
Differentiate wrt α_i and get gradient:

$$\nabla\psi = [\lambda_1\sin\alpha_1, \lambda_2\sin\alpha_2 - \lambda_3\cos\alpha_2]^{\mathsf{T}}.$$

Leave it as it is! We are done!

FINDING $\nabla_\alpha \mathcal{C}(x_N)$ WITHOUT INTEGRATING ADJOINT

- We illustrate with Euler's rule. Result is however general.

- To find $\nabla_\alpha \mathcal{C}(x_N)$, multiplier approach is welcome; look at x_n and slopes k_n as intermediate variables γ_i. The Lagrangian is

$$\mathcal{C}(x_N) - \lambda_0^{\mathsf{T}}(x_0 - \alpha) - \sum_{n=0}^{N-1} h_n \lambda_{n+1}^{\mathsf{T}} \left[\frac{1}{h_n}(x_{n+1} - x_n) - k_n \right]$$
$$- \sum_{n=0}^{N-1} h_n \Lambda_n^{\mathsf{T}} \Big[k_{n,i} - f(x_n, t_n) \Big].$$

(Note this may be regarded as the result of using a quadrature rule on continuous Lagrangian.)

MATHS ARE MAGICAL

- Use of the Lagrange multiplier recipe results in:

$$\begin{aligned} x_0 &= \alpha, \\ x_{n+1} &= x_n + h_n f(x_n, t_n), \\ \lambda_{n+1} &= \lambda_n - h_n \partial_x f(x_n, t_n)^{\mathsf{T}} \lambda_{n+1}, \\ \nabla_x \mathcal{C}(x_N) &= \lambda_N, \\ \nabla_\alpha \mathcal{C}(x_N) &= \lambda_0. \end{aligned}$$

- But these are just the equations that we used to find $\nabla_\alpha \mathcal{C}(x_N)$ with the clever PRK integrator.

- Standard chain rule implies a (hidden) application of symplectic PRK.

VARIATIONAL–ADJOINT/δ–λ/FORWARD–BACKWARD

• λ_N, λ_{N-1}, λ_{N-2}, . . . successively yield $\nabla_{x_N}\mathcal{C}(x_N)$, $\nabla_{x_{N-1}}\mathcal{C}(x_N)$, $\nabla_{x_{N-2}}\mathcal{C}(x_N)$, . . . (backward propagation of the gradient).

• When applying the chain rule we have to multiply Jacobian matrices; say $y = y(z)$, $y = y(x)$, $w = w(x)$.

$$\frac{\partial(w)}{\partial(x)}\frac{\partial(x)}{\partial(y)}\frac{\partial(y)}{\partial(z)} \qquad \text{that we write} \qquad J_3J_2J_1.$$

Choice of ordering $J_1 \to J_2J_1 \to J_3J_2J_1$ (forward, natural?) versus $J_3 \to J_3J_2 \to J_3J_2J_1$ (backward) is important.

• Say J_3 is 3×200, J_2 is 200×100 and J_1 is 100×90 (as in computing differential of map from $\mathbb{R}^{90}$ to $\mathbb{R}^3$ if there are $100 + 200$ intermediate variables . . .)

IIII. OPTIMAL CONTROL

CONTROL PROBLEMS

Choose control function $u(t)$ so as to minimise a cost function $\mathcal{C}(x(t_0+T))$, subject to $(d/dt)x = f(x,u,t)$ and $x(t_0) = \alpha$.

With the same ideas and techniques we may deal with:

* Constrained controls (say $u > 0$).

* Cost functionals:

$$\mathcal{C}(x(t_0+T)) + \int_{t_0}^{t_0+T} \mathcal{D}(x(t),u(t),t)\,dt.$$

* . . . . . .

OPTIMALITY CONDITIONS

- Variational equation (ζ is the variation in u):

$$\frac{d}{dt}\delta = \partial_x f(x(t), u(t), t)\,\delta + \partial_u f(x(t), u(t), t)\,\zeta.$$

- Adjoint system and control constraints are:

$$\frac{d}{dt}\lambda = -\partial_x f(x(t), u(t), t)^{\mathsf{T}}\,\lambda, \;\; \partial_u f(x(t), u(t), t)^{\mathsf{T}}\lambda(t) = 0.$$

- These ensure $\lambda(t_0+T)^{\mathsf{T}}\delta(t_0+T) = \lambda(t_0)^{\mathsf{T}}\delta(t_0)$. Now $\delta(t_0) = 0$, if we impose final condition $\lambda(t_0 + T) = \nabla\mathcal{C}(x(t_0 + T))$, we get $\nabla\mathcal{C}(x(t_0+T))^{\mathsf{T}}\delta(t_0+T) = 0$ (first-order condition for optimality).

- [Technical assumption: constraints define control u as functions of dual variable λ.]

DISCRETISATION

• Discretise $(d/dt)x = f(x, u, t)$ by RK so that control problem becomes an optimization problem in finite-dimensional space.

• eg for Euler's rule: $x_0 = \alpha$, $x_{n+1} = x_n + h_n f(x_n, u_n, t_n)$, $n = 0, \ldots, N-1$.

• $\mathcal{C}(x_N)$ is a function of state variables and controls and the RK equations act as constraints. We have to impose conditions to determine discrete states and controls to achieve optimality.

FINDING THE OPTIMAL DISCRETE SOLUTION

• Indirect approach. Numerically integrating the adjoint system with RK method used for primal system will NOT guarantee that discrete solution satisfies necessary first-order condition to minimize cost.

• For that to happen, method has to be symplectic (or adjoint system has to be integrated to get the clever overall symplectic PRK method we saw above).

• Direct approach. If we ignore the adjoint system and just minimize the cost as a function constrained by the RK equations, then we automatically arrive at the same clever symplectic PRK discretisation (Hager 2000). (The magic of maths.)

V. APPLICATION TO MECHANICS

LAGRANGIAN MECHANICS

• May be derived from Hamilton's principle (1835): impose end conditions $x(t_0) = \alpha$, $x(t_0 + T) = \beta$, and make stationary the action

$$\int_{t_0}^{t_0+T} \mathcal{L}(x(t), u(t), t)\, dt$$

(x are the Lagrangian co-ordinates and u their velocities).

• This is a control problem for the system $(d/dt)x = u$ with controls u. The dual variables λ turn out to be the momenta conjugated to x.

• Application of the theory above to the present instance just yields the derivation of symplectic PRK algorithms as variational integrators (Suris 1990).