Backward Error Analysis and Applications

Nicky Cordua Mattsson

Department of Mathematics and Computer Science
Office: Saunaen

May 6, 2015

Table of contents

Introduction

Consider the differential equation:

$$\dot{y} = f(y) \tag{0.1}$$

and a numerical method $\Phi_h(y)$, which produces the approximations $y_0, y_1, \ldots, y_n$. Then a forward error analysis consists of the study of the errors $y_1 - \varphi_h(y_0)$ (local error) and $y_n - \varphi_{nh}(y_0)$ (global error).

In Backward analysis we search for a *modified differential equation* $\dot{\tilde{y}} = f_h(\tilde{y})$ of the form

$$\dot{\tilde{y}} = f(\tilde{y}) + \sum_{i=2} h^{i-1} f_i(\tilde{y}) \tag{0.2}$$

such that $y_n = \tilde{y}(nh)$. Note that the polynomial does in general diverge and one needs to truncate it properly. Further note that we have $y_n - y(nh) = \tilde{y}(nh) - y(nh)$ so we want to study the difference of the vectorfields.

Example

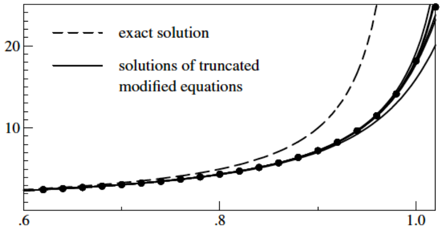
Consider the scalar IVP:

$$\dot{y} = y^2, y(0) = 1 \tag{0.3}$$

making a Taylor expansion of $\tilde{y}(t+h)$ and compare with the numerical solution, we obtain that the modified differential equation:

$$\dot{\tilde{y}} = \tilde{y}^2 - h\tilde{y}^3 + h^2\frac{3}{2}\tilde{y}^4 - h^3\frac{8}{3}\tilde{y}^5 \pm \dots \tag{0.4}$$

and the figure:



Modified Differential Equation

Theorem

Suppose that the method $y_{n+1} = \Phi_h(y_n)$ is of order p, i.e.

$$\Phi = \varphi_h(y) + h^{p+1}\delta_{p+1}(y) + \mathcal{O}(h^{p+2}) \tag{1.1}$$

where $\varphi_t(y)$ is the exact flow of $\dot{y} = f(y)$, and $h^{p+1}\delta_{p+1}$ is the leading term of the local truncation error. The modified equation then satisfies:

$$\dot{\tilde{y}} = f(\tilde{y}) + h^p\delta_{p+1}(\tilde{y}) + \sum_{i=p+2} h^{i-1} f_i(\tilde{y}), \quad \tilde{y}(0) = y_0 \tag{1.2}$$

Modified Equations of Symmetric Methods

Theorem (Adjoint Methods)

Let $f_j(y)$ be the coefficient functions of the modified equation for the method Φ_h. Then, the coefficient functions $f_j^*(y)$ of the modified equation for the adjoint method Φ_h^* satisfy:

$$f_j^*(y) = (-1)^{j+1} f_j(y) \tag{1.3}$$

Theorem (Symmetric methods)

The coefficient functions of the modified equation of a symmetric method satisfies that $f_j = 0$ whenever j is even. So that the equation has an expansion of even powers.

Theorem (Reversible Methods)

Consider a ρ-reversible differential equation and numerical method. Then every truncation of the modified differential equation is again ρ-reversible.

Modified Equations of Symplectic Methods

Theorem

If a symplectic method method $\Phi_h(y)$ is applied to a Hamiltonian system with a smooth Hamiltonian $H : \mathbb{R}^{2d} \to \mathbb{R}$, then the modified equation is also Hamiltonian.

Theorem

Assume that the symplectic method Φ_h has a generating function

$$S(P,q,h) = hS_1(P,q) + h^2 S_2(P,q) + h^3 S_2(P,q) + \dots \quad (1.4)$$

with smooth $S_j(P,q)$ defined on an open set D. Then, the modified differential equation is a Hamiltonian system with

$$\tilde{H}(p,q) = H(p,q) + hH_2(p,q) + \dots \quad (1.5)$$

where the Hamiltonians are defined and smooth on the whole of D.

Modified Equations of Symplectic Methods

Theorem

A symplectic (partitioned) Runge-Kutta method applied to a system with smooth Hamiltonian $H : D \to \mathbb{R}$ with $D \subset \mathbb{R}^{2d}$ has a modified Hamiltonian (as above) with smooth Hamiltonians from D to $\mathbb{R}$.

Modified Equations of Poisson Integrators

Theorem

If a Poisson integrator $\Phi_h(y)$ is applied to the Poisson system, then the modified equation is locally a Poisson system.

Theorem

If $H(y)$ and $B(Y)$ are defined and smooth on a simply connected domain D, and if $B(y)$ is invertible on D, then a Poisson integrator has a modified equation with smooth Hamiltonians defined on D

Modified Equation of Methods on Manifolds

Theorem

Let $\Phi_h : \mathcal{M} \to \mathcal{M}$ be an integrator on the manifold $\mathcal{M}$, with Φ_h depending smoothly on (y, h). Then, there exists a modified differential equation on the manifold, such that:

$$\dot{\tilde{y}} = f(\tilde{y}) + hf_2(\tilde{y}) + \dots \tag{1.6}$$

with smooth $f_j(y) \in T_y\mathcal{M}$ such that $\varphi_{r,h}(y) = \Phi_h(y) + \mathcal{O}(h^{r+1})$

Theorem

Let the integrator $\Phi_h : U \to \mathbb{R}^n$ be differentiable in h, and let $\mathcal{M} = \{y \in U; g(y) = 0\}$ with differentiable $g : U \to \mathbb{R}^m$. If the coefficient functions $f_j(y)$ of the modified differential equation satisfy $g'(y)f_j(y) = 0$ for all j and all $y \in \mathcal{M}$, then the restriction of Φ_h to $\mathcal{M}$ defines an integrator on $\mathcal{M}$.

Modified Equation of Methods on Manifolds

Theorem

Consider a differential equation $\dot{y} = f(y)$ with invariant $I(y)$. If the numerical method preserves this invariant, then every truncation of the modified equation has $I(y)$ as invariant.

Theorem

Consider a differential equation $\dot{y} = f(y)$ with invariant $I(y)$. If the numerical method, Φ_h, is differentiable in h, and if every truncation of the modified equation has $I(y)$ as invariant, then the numerical method preserves the invariant $I(y)$ exactly.

Modified Equation of Constraint Hamiltonian Systems

Theorem

A symplectic integrator $\Phi_h : \mathcal{M} \to \mathcal{M}$ for the constrained Hamiltonian system has a modified equation which is locally a constrained Hamiltonian system.

Modified Equations for Variable Step Sizes

Problem: The modified differential equation of the numerical integrator depends on the step size. So if the stepsize changes, so does the modified equation. This causes bad long time behaviour

Solution: Using prior adaptive step size approach, in stead of the posterior, as we have seen in the previous presentation. Here we apply a fixed step size method to a transformed system, making the modified differential equation the same for all steps. Explaining the good long time behaviour.

Rigorous Estimates - Local Error

So far we have not been taking care of convergence, however in the following we will:

- Sketch that even in simple situations it does not converge
- Give bounds on the coefficient functions, and find an optimal truncation index
- Determine the local error of the numerical solution of $\dot{y} = f(t, y)$ and the exact solution of the truncated modified equation

Rigorous Estimates - Local Error

Consider the differential equation:

$$\dot{y} = f(t), \quad y(0) = 0 \tag{2.1}$$

to which we apply the trapezoidal rule:
$y_1 = \frac{h}{2}(f(t) + f(t+h))$, we have the expansion:

$$\begin{aligned} \Phi_h(t, y) &= y + \frac{h}{2}(f(t) + f(t+h)) \\ &= y + hf(t) + \frac{h^2}{2}f'(t) + \frac{h^3}{4}f''(t) + \dots \end{aligned}$$

Rigorous Estimates - Local Error

The modified equation is necessarily of the form:

$$\dot{\tilde{y}} = f(t) + hb_1 f'(t) + h^2 b_2 f''(t) \ldots \tag{2.2}$$

The coefficients b_i of the modified equation can be computed by setting $f(t) = e^t$ and using the relation $\Phi_h(t, y) = \tilde{y}(t+h)$ with initial value $\tilde{y}(t) = y$ yields:

$$\frac{h}{2}(e^h + 1) = (1 + b_1 h + b_2 h^2 + \ldots)(e^h - 1) \tag{2.3}$$

which shows that $b_1 = 0$ and $b_k = B_k/k! \approx C(2\pi)^{-k}$ where B_k are Bernoulli numbers, and the series diverge as soon as $f^{(k)}(t) \geq k! M R^{-k}$

Rigorous Estimates - Local Error

Definition

Define the complex ball $B_{2R}(y_0)$ as the following:

$$B_{2R}(y_0) := \{y \in \mathbb{C}; \|y - y_0\| \leq 2R\} \tag{2.4}$$

In the following assume that:

$$\forall y \in B_{2R} : \|f(y)\| \leq M \tag{2.5}$$

and that we have the following Taylor expansion:

$$\Phi_h(y) = y + hf(y) + h^2 d_2(y) + h^3 d_3(y) + \ldots \tag{2.6}$$

Bound on the Numerical Solution

Theorem

For a Runge-Kutta method let

$$\mu = \sum_{i=1}^{s} |b_i|, \quad \kappa = \max_{i=1,\dots,s} \sum_{j=1}^{s} |a_{ij}| \tag{2.7}$$

If $f(y)$ is differentiable in the complex ball $B_{2R}(y_0)$ and satisfies the assumption, then the coefficient functions $d_j(y)$ are differentiable in $B_R(y_0)$ and satisfy:

$$\|d_j(y)\| \leq \mu M \left(\frac{2\kappa M}{R}\right)^{j-1}, \quad \|y - y_0\| \leq R \tag{2.8}$$

Note: The first order condition (called the consistency condition) i.e. $\sum_{i=1}^{s} b_i = 1$ causes methods with only positive weights to all have $\mu = 1$

Bound on the Modified Differential Equation

Lemma

If the numerical method has an expansion of the previously seen form, then the function $f_j(y)$ of the modified differential equation satisfy

$$f_j(y) = d_j(y) - \sum_{i=2}^{j} \frac{1}{i!} \sum_{k_1+\ldots+k_i=j} \left(D_{k_1} \ldots D_{k_{i-1}} f_{k_i}\right)(y) \quad (2.9)$$

where $k_m \geq 1$ for all m. Observe that the right-hand expression only involves $f_k(y)$ with $k < j$.

Bound on the Modified Differential Equation

Theorem

Let $f(y)$ be analytic in $B_{2R}(y_0)$, let the Taylor series coefficients of the numerical method be differentiable in $B_R(y_0)$, and assume that the assumptions be fulfilled. Then we can bound the coefficient functions of the modified differential equation.

$$\|f_j(y)\| \leq \ln(2)\eta M \left(\frac{\eta M j}{R}\right)^{j-1}, \quad y \in B_{R/2}(y_0) \qquad (2.10)$$

where $\eta = 2\max(\kappa, \mu/(2\ln(2) - 1))$

Choice of Truncation

Consider the following:

$$\dot{\tilde{y}} = F_N(\tilde{y}), \quad F_N(\tilde{y}) = f(\tilde{y}) + hf_2(\tilde{y}) + \ldots + h^{N-1} f_N(\tilde{y}) \tag{2.11}$$

with initial value $\tilde{y}(0) = y_0$. It is common in the theory of asymptotic expansions to truncate the series where the corresponding term is the smallest. Then from the previous bounds we assume that the truncation index N satisfies :

$$N \leq \frac{h_0}{h}, \quad h_0 = \frac{R}{e\eta M}. \tag{2.12}$$

Then for a pth order method we obtain:

$$\|F_N(y) - f(y)\| \leq cMh^p \tag{2.13}$$

Choice of Truncation

Theorem

Let $f(y)$ be differentiable in $B_{2R}(y_0)$, let the coefficients $d_j(y)$ be differentiable in $B_R(y_0)$, and assume that the previous assumptions hold. If $h \leq h_0/4$, then there exists $N = N(h)$ (largest integer fulfilling Equation (2.12)) such that the difference between the numerical solution and the exact solution of the truncated modified equation is bounded:

$$\|\Phi_h(y_0) - \tilde{\varphi}_{N,h}(y_0)\| \leq h\gamma M e^{-h_0/h} \qquad (2.14)$$

where $\gamma = e(2 + 1.65\eta + \mu)$ depends only on the method.

Theorem (Chartier, Faou and Murua 2005)

The only symplectic method (as B-series) that conserves the Hamiltonian for arbitrary $H(y)$ is the exact flow of the differential equation.

Introductory Things

Definition (Completely Integrable)

A Hamiltonian system with Hamiltonian $H : M \to \mathbb{R}$ (M an open subset of $\mathbb{R}^d \times \mathbb{R}^d$) is called completely integrable if there exist smooth functions $F_1, F_2, \dots, F_d : M \to \mathbb{R}$ with the following properties:

1. $F_1 = H$
2. $F_1, \dots, F_d$ *are in involution (i.e. all* $\{F_i, F_j\} = 0$*) on* M
3. *The gradients of* $F_1, \dots, F_d$ *are linearly independent at every point of* M
4. *The solution trajectories of the Hamiltonian systems with Hamiltonian* F_i *exist for all times and remain in* M*.*

Definition (Level Set)

For $x = (x_i) \in \mathbb{R}^d$ *we define the level set*

$$M_x = \{(p, q) \in M; F_i(p, q) = x_i, \text{ for } i = 1, \dots, d\} \quad (3.1)$$

Introductory Things

Theorem (Arnold † 2010, Liouville †1882)

Let $F_1, \ldots, F_d : M \to \mathbb{R}$ be invariants of a completely integrable system as in the previous definition. Suppose that the level sets M_x are compact and connected for all x in a neighbourhood of $x_0 \in \mathbb{R}^d$.
Then there are neighbourhoods B of x_0 and D of 0 in $\mathbb{R}^d$ such that the following holds:
-For every $x \in B$, the level set M_x is a d-dimensional torus that is invariant under the flow of the system with Hamiltonian F_i
-There exists a bijective symplectic transformation

$$\psi : D \times \mathbb{T}^d \to \bigcup_{x \in B} M_x \subset \mathbb{R}^d \times \mathbb{R}^d : (a, \theta) \mapsto (p, q) \quad (3.2)$$

such that $(F_i \circ \psi)(a, \theta)$ depends only on a, further the variables (a, θ) are called action variables.

Introductory Things

Remark 1: If the level sets M_x are not compact, then the proof of the theorem shows that M_x is diffeomorphic to a Cartesian product of of circles and straight lines.

Remark 2: If the level sets M_x are not compact, then the proof of the theorem shows that still the invariants only depends on a.

Remark 3: If the Hamiltonian is differentiable, then the proof shows that also the transformation is differentiable.

Remark 4: In case of a completely integrable Hamiltonian $H = F_1$, then the action variables, i.e. a_i, are constant

Definition (Siegel's Diophantine Condition)

Let γ and ν be real positive constants, ω the frequencies in the new coordinates and $k \in \mathbb{Z}^d$, $k \neq 0$, such that $|k| = \sum_i |k_i|$. Then we call the following equation for Siegel's Diophantine condition:

$$|k \cdot \omega| \geq \gamma |k|^{-\nu} \tag{3.3}$$

Linear Error Growth and Near Preservation of First Integrals

Consider a completely integrable Hamiltonian system:

$$\dot{p} = \frac{\partial H}{\partial q}(p, q), \quad \dot{q} = \frac{\partial H}{\partial p}(p, q) \tag{3.4}$$

Apply a symplectic method to obtain a solution sequence. We assume that the Hamiltonian is differentiable and that the conditions of the Arnold-Liouville theorem are fulfilled.

Linear Error Growth and Near Preservation of First Integrals

Make a change of coordinates to action-angle coordinates $(p, q) = \psi(a, \theta)$, denote the inverse transformation:

$$(a, \theta) = (I(p, q), \Theta(p, q)) \tag{3.5}$$

Recall: Completely integrable system, then a is constant, implying that all I_i are invariants of the system.
In the action-angle system denote the Hamiltonian as $\mathcal{H}(a) = H(p, q)$, further denote the frequencies:

$$\omega(a) = \frac{\partial \mathcal{H}}{\partial a}(a) \tag{3.6}$$

which we consider in a neighbourhood of some $a^* \in \mathbb{R}^d$

Linear Error Growth and Near Preservation of First Integrals

Theorem

Consider applying a symplectic numerical integrator of order p to the completely integrable Hamiltonian system. Suppose that $\omega(a^*)$ satisfy the Diophantine condition.
Then, there exist positive constants C, c and h_0 such that the following holds for all step sizes $h \leq h_0$: Every numerical solution starting with $\|I(p_0, q_0) - a^*\| \leq c|log(h)|^{-\nu-1}$ satisfies:

$$\begin{aligned} \|(p_n, q_n) - (p(t), q(t))\| &\leq Cth^p, \quad t \leq h^{-p} \\ \|I(p_n, q_n) - I(p(t), q(t))\| &\leq Cth^p \end{aligned}$$

The constants h_0, c, C depend on d, γ, ν, on bounds of the differentiable H, on a complex neighbourhood of the torus and on the numerical method.

Linear Error Growth and Near Preservation of First Integrals

Remark 4: The linear error growth holds also when the symplectic method is applied to a perturbed integrable system with a perturbation parameter ϵ bounded by a positive power of the step size: $\epsilon \leq Kh^{\alpha}$ for some $\alpha > 0$.

Near-Invariant Tori on Exponential Long Times

"Perturbation theory if in fact an outgrowth of the necessity to determine the orbits with ever greater accuracy. This problem can be solved today, but in what is for the theoretician a rather disappointing way. With modern calculating machines, one is now able to compute directly results even more accurately then those provided by perturbation theory" - J. Moser 1978.

Idea: Use perturbation theory to compare the original differential equation to the modified differential equation.

One can now bound the perturbation terms first locally, then on exponentially long times. (Comparison of Fourier expansion and power expansion of the perturbation term)

Kolmogorov († 1987), Arnold († 2010) and Moser († 1999)[1] (KAM Theory)

[1] Russian Mathematicians

Near-Invariant Tori on Exponential Long Times

Theorem

Consider applying a symplectic numerical integrator of order p to the completely integrable and differentiable Hamiltonian system. Suppose that $\omega(a^*)$ satisfies the Diophantine condition.

Then there exist positive constants c_0, c, C and h_0 such that the following holds for all step sizes $h \leq h_0$ and for all $\mu \leq \min(p, \alpha)$ with $\alpha = \nu + d + 1$: every numerical solution starting with $\|I(p_0, q_0) - a^*\| \leq c_0 h^{2\mu}$ satisfies:

$$\|I(p_n, q_n) - I(p_0, q_0)\| \leq Ch^p, \quad nh \leq exp(ch^{-\mu/\alpha}) \quad (4.1)$$

The constants has same dependencies as in the previous theorem.

Kolmogorov's Theorem on Invariant Tori

Theorem (Kolmogorov's Iteration 1954)

Consider a differentiable Hamiltonian $H(a, \theta)$, defined for a in a neighbourhood of $0 \in \mathbb{R}^d$ and $\theta \in \mathbb{T}^d$, for which the linearisation does not depend on the angles:

$$H(a, \theta) = c + \omega a + \frac{1}{2} a^T M(a, \theta) a. \tag{4.2}$$

Suppose that $\omega \in \mathbb{R}^d$ satisfies the Diophantine condition, and that the angular average $\bar{M}_0$ of $M(0, \cdot)$ is an invertible matrix with:

$$\|\bar{M}_0 v\| \geq \mu \|v\| \tag{4.3}$$

with positive constants γ, ν, μ. Let $H_\varepsilon(a, \theta) = H(a, \theta) + \varepsilon G(a, \theta)$ be a small differentiable perturbation of the original system. Then there is a differentiable symplectic transformation, which takes the size of the perturbation from $O(\varepsilon)$ to $O(\varepsilon^2)$

Kolmogorov's Theorem on Invariant Tori

Theorem (Kolmogorov's theorem, Arnold 1963 and later Benettin, Galgani, Giorgilli & Strelcyn 1984)

Consider a differentiable Hamiltonian $H(a, \theta)$ *in the above described case. Again we let* $H_\varepsilon(a, \theta) = H(a, \theta) + \varepsilon G(a, \theta)$ *be a differentiable perturbation of the system.*
Then there is a differentiable symplectic transformation $\phi_\varepsilon : (b, \varphi) \mapsto (a, \theta)$, $O(\varepsilon)$ *close to the identity and depending analytically on* ε, *which puts the perturbed Hamiltonian back to the form:*

$$H_\varepsilon(a, \theta) = c_\varepsilon + \omega b + \frac{1}{2} b^T M_\epsilon(b, \varphi) b \qquad (4.4)$$

The system therefore has the invariant torus (KAM Tori) $\mathcal{T}_\varepsilon = \{b = 0, \varphi \in \mathbb{T}^d\}$ *carrying a quasi-periodic flow with the same frequencies* ω *as the unperturbed system.*

KAM Tori under Symplectic Discretisation

Consider a Hamiltonian system in the case where we fulfil the conditions for Kolmogorov's theorem, and apply a numerical method to the perturbed system, then the modified equation is:

$$\tilde{\mathcal{H}}(p,q) = \omega b + \frac{1}{2} b^T M_\varepsilon(b,\varphi) b + h^p \tilde{G}(b,\varphi) \tag{4.5}$$

Now apply Kolmogorov's theorem again, then we obtain:

$$\tilde{\mathcal{H}}(p,q) = \omega c + \frac{1}{2} c^T M_{\varepsilon,h}(c,\psi) c \tag{4.6}$$

combining this with the earlier bound between the exact solution of the modified equation and the numerical method yields the following theorem.

KAM Tori under Symplectic Discretisation

Theorem (Hairer and Lubich 1997)

In the above situation for a symplectic integrator of order p used with sufficiently small step size h, there is a modified Hamiltonian $\tilde{\mathcal{H}}(p,q)$ with an invariant KAM torus of the original (perturbed) Hamiltonian $\mathcal{H}$, such that the difference between any numerical solution starting on the torus and the solution of the modified Hamiltonian system with the same starting values can be bounded by:

$$\|(p_n, q_n) - (\tilde{p}(t), \tilde{q}(t))\| \leq Ce^{-\kappa/h}, \quad t \leq e^{\kappa/h} \qquad (4.7)$$

The constants C and κ are independent of n, h and ε.

One can impose stronger demands on the ω to get the near conservation in general.

Invariant Tori of Symplectic Maps

Lemma

Suppose that $\omega \in \mathbb{R}^2$ satisfies the Diophantine condition, and let $h_0 > 0$. For any choice of positive $\gamma^* \leq \gamma$ and $\nu^* > \nu + d + r$ with $r > 1$, the set:

$$Z(h_0) = \{h \in (0, h_0); h \text{ does not satisfy the stronger demand}\} \tag{4.8}$$

has Lebesgue measure:

$$m(Z(h_0)) \leq C \frac{\gamma^*}{\gamma} h_0^{r+1}. \tag{4.9}$$

Note that for $h_0 \to 0$ then the probability of choosing an h which makes the Torus near invariant for all times, goes to 1.

Reversible Perturbation Theory

The previous can very easily be extended to symmetric methods applied to reversible systems.

"The results an techniques are largely analogous to those of the previous chapter - the extent of the analogy may indeed be seen as the most surprising feature of this chapter" - E. Hairer, G. Wanner and C. Lubich.