Regularized dynamical parametric approximation

Christian Lubich
Univ. Tübingen

Talk based on joint work with
Caroline Lasser, Michael Feischl, Jörg Nick

arXiv:2403.19234

Exploiting Algebraic and Geometric Structure in Time-Integration Methods,
Scuola Normale Superiore, Pisa, 4 April 2024

Outline

Outline

Nonlinear parametric approximation of evolution

High-dimensional ODE or PDE

$$\dot{y} = f(y), \qquad y(0) = y_0$$

to be approximated via nonlinear parametrization

$$y(t) \approx u(t) = \Phi(q(t))$$

with unknown time-dependent parameters $q(t) \in \mathcal{Q}$

Our case of interest here is an irregular parametrization:
$\Phi'(q)$ has arbitrarily small singular values and possibly varying rank.

Loss of geometric structure

Motivation from applications

Parametrization by

- Multiple Gaussians
- Tensor networks
- Deep neural networks

$\Phi'(q)$ typically has numerous very small singular values.

Dynamical parametric approximation

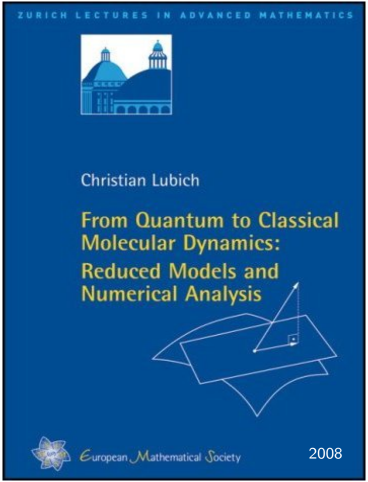
How to choose the time-dependent parameters $q(t)$?

Minimum defect condition: Given $q(t)$ and $u(t) = \Phi(q(t))$, determine $\dot{q}(t)$ and $\dot{u}(t) = \Phi'(q(t))\,\dot{q}(t)$ such that

$$\|\dot{u}(t) - f(u(t))\| = \min!$$

Widely used in quantum physics / chemistry:
Dirac–Frenkel time-dependent variational principle (Dirac 1930)
rich geometry

Tangent space projection



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

is approximated by

$$\dot{u} = P_u f(u)$$

---

in quantum physics/chemistry:
Dirac–Frenkel time-dependent variational principle

Dynamical parametric approximation

The least-squares problem for $\dot{q}$

$$\|\dot{u} - f(u)\| = \|\Phi'(q)\,\dot{q} - f(\Phi(q))\| = \min!$$

is ill-posed when $\Phi'(q)$ has small singular values.

What can be done when the geometry breaks down?

Regularized dynamical parametric approximation

Given $q(t)$ and $u(t) = \Phi(q(t))$,
determine $\dot{q}(t)$ and $\dot{u}(t) = \Phi'(q(t))\,\dot{q}(t)$ from the
regularized linear least-squares problem

$$\delta(t)^2 := \|\dot{u}(t) - f(u(t))\|^2 + \varepsilon^2 \|\dot{q}(t)\|_{\mathcal{Q}}^2 = \min!$$

$\varepsilon > 0$ small regularization parameter (possibly state-dependent)

This gives a differential equation for the parameters $q(t)$, which is solved numerically by standard time integrators.

Cf. Ritcheson et al. (2015) for multi-Gaussians in quantum dynamics.

Outline

Ill-posedness of the regularized problem

Consider the sensitivity to initial values:

Let $q(t)$, $\widetilde{q}(t)$ be the solutions to initial values $q(0)$, $\widetilde{q}(0)$, and $u(t) = \Phi(q(t))$, $\widetilde{u}(t) = \Phi(\widetilde{q}(t))$

Perturbation bound:

$$\begin{aligned} &\|u(t) - \widetilde{u}(t)\| + \varepsilon \|q(t) - \widetilde{q}(t)\|_{\mathcal{Q}} \\ &\qquad \le e^{\omega t} \left(\|u(0) - \widetilde{u}(0)\| + \varepsilon \|q(0) - \widetilde{q}(0)\|_{\mathcal{Q}} \right) \end{aligned}$$

with

$$\omega = L + \tfrac{5}{2} \frac{\beta}{\varepsilon} \frac{\bar{\delta}}{\varepsilon} \sim \frac{\beta}{\varepsilon} \quad \text{when} \quad \beta \neq 0,$$

where β is an upper bound of Φ'',
L is a local Lipschitz constant of f,
$\bar{\delta}$ is an upper bound of the defect size

Severely ill-posed for nonlinear parametrizations

A basic principle of numerical analysis

Avoid solving ill-posed subproblems when you aim for a stable algorithm for a well-posed problem.

A basic principle of numerical analysis

Avoid solving ill-posed subproblems when you aim for a stable algorithm for a well-posed problem.

We ignore this good advice.

Well-posedness up to the defect size

$$\delta(t)^2 := \|\dot{u}(t) - f(u(t))\|^2 + \varepsilon^2 \|\dot{q}(t)\|_{\mathcal{Q}}^2 = \text{min!}$$

$$u(t) = \Phi(q(t)), \quad \widetilde{u}(t) = \Phi(\widetilde{q}(t))$$

Perturbation bound:

$$\|u(t) - \widetilde{u}(t)\| \le e^{\ell t}\|u(0) - \widetilde{u}(0)\| + \int_0^t e^{\ell(t-s)} \left(\delta(s) + \widetilde{\delta}(s)\right) ds$$

with the one-sided (local) Lipschitz constant ℓ :

$$\langle y - \widetilde{y}, f(y) - f(\widetilde{y}) \rangle \le \ell \, \|y - \widetilde{y}\|^2 \qquad \forall y, \widetilde{y}$$

Proof

$$\dot{u} = f(u) + d \quad \textit{with} \quad \|d\| \le \delta$$
$$\dot{\widetilde{u}} = f(\widetilde{u}) + \widetilde{d} \quad \textit{with} \quad \|\widetilde{d}\| \le \widetilde{\delta}$$

Subtract, take inner product with $u - \widetilde{u}$.
On the left:

$$\frac{1}{2}\frac{d}{dt}\|u - \widetilde{u}\|^2 = \|u - \widetilde{u}\| \cdot \frac{d}{dt}\|u - \widetilde{u}\|$$

On the right: use the one-sided Lipschitz condition and Cauchy–Schwarz to bound by

$$\le \|u - \widetilde{u}\| \left(\ell\|u - \widetilde{u}\| + \|d - \widetilde{d}\|\right)$$

Use the rough bound

$$\|d - \widetilde{d}\| \le \delta + \widetilde{\delta},$$

divide both sides by $\|u - \widetilde{u}\|$ and use Gronwall's inequality. □

---

simple, uses standard arguments

Well-posedness up to the defect size

$$\delta(t)^2 := \|\dot{u}(t) - f(u(t))\|^2 + \varepsilon^2 \|\dot{q}(t)\|_{\mathcal{Q}}^2 = \text{min!}$$

$$u(t) = \Phi(q(t)), \quad \widetilde{u}(t) = \Phi(\widetilde{q}(t))$$

Perturbation bound:

$$\|u(t) - \widetilde{u}(t)\| \leq e^{\ell t} \|u(0) - \widetilde{u}(0)\| + \int_0^t e^{\ell(t-s)} \left(\delta(s) + \widetilde{\delta}(s)\right) ds$$

with the one-sided (local) Lipschitz constant ℓ :

$$\langle y - \widetilde{y}, f(y) - f(\widetilde{y}) \rangle \leq \ell \, \|y - \widetilde{y}\|^2 \qquad \forall y, \widetilde{y}$$

A posteriori error bound

- $\dot{y} = f(y), \qquad y(0) = u(0) = \Phi(q(0))$
- Regularized parametric approximation $u(t) = \Phi(q(t))$ with
$$\delta(t)^2 = \|\dot{u}(t) - f(u(t))\|^2 + \varepsilon^2 \|\dot{q}(t)\|_{\mathcal{Q}}^2 \quad \text{minimal}$$

Error bound:

$$\|u(t) - y(t)\| \le \int_0^t e^{\ell(t-s)}\, \delta(s)\, ds$$

simple proof as before

A priori error bound

We want to bound the error of the algorithm in terms of approximability of the solution (cf. Céa's Lemma).

Pointwise approximability of the solution $y(t)$ is *not* sufficient here.

Error bounds require approximability of the solution derivative $\dot{y}(t)$ in the tangent spaces at all $u = \Phi(q)$ close to $y(t)$.

(No details in this talk, see our paper.)

Outline

Regularized Euler time-stepping

$$q_{n+1} = q_n + h\dot{q}_n, \qquad u_{n+1} = \Phi(q_n),$$

where $\dot{q}_n$ solves the regularized linear least-squares problem

$$\delta_n^2 := \|\Phi'(q_n)\dot{q}_n - f(u_n)\|^2 + \varepsilon_n^2\,\|\dot{q}_n\|^2 = \min!$$

and $h > 0$ is the stepsize.

Local error bound

Under the stepsize restriction

$$h\delta_0 \leq c\varepsilon_0^2$$

the error after one step starting from $y_0 = u_0 = \Phi(q_0)$
is bounded by

$$\|u_1 - y(t_1)\| \leq c_1 h\delta_0 + c_2 h^2$$

with $c_1 = 1 + c\beta$ and $c_2 = \frac{1}{2} \max_{t_0 \leq t \leq t_1} \|\ddot{y}(t)\|^2$,
where β is a bound of Φ''.

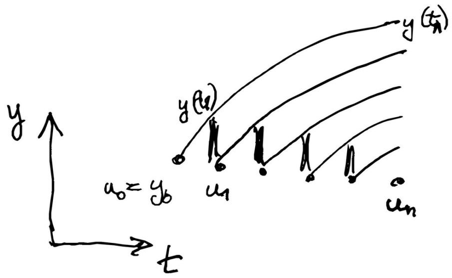
The stepsize restriction yields $\|\Phi''(q)[q_1 - q_0, q_1 - q_0]\|_{\mathcal{Q}}^2 \leq c\beta\, h\delta_0$.

Global error bound

Under the stepsize restriction $h_n \delta_n \le c\varepsilon_n^2$, the error is bounded by

$$\|u_n - y(t_n)\| = O(\delta + h), \qquad 0 \le t_n \le \bar{t}.$$

This is obtained from the local error bound via
Lady Windermere's fan with error propagation by the exact flow:
(see numerical ODE book by Hairer, Nørsett & Wanner 1987)



Regularized Runge-Kutta time discretization

Explicit Runge-Kutta method of order p applied to the differential equation for the parameters q:
Compute internal stages (for $i = 1, \dots, s$)

$$q_{n,i} = q_n + h \sum_{j=1}^{i-1} a_{ij} \dot{q}_{n,j}, \qquad u_{n,i} = \Phi(q_{n,i}),$$

and $\dot{q}_{n,i}$ solving the regularized linear least squares problem

$$\delta_{n,i}^2 := \|\Phi'(q_{n,i})\dot{q}_{n,i} - f(u_{n,i})\|^2 + \varepsilon_n^2 \|\dot{q}_{n,i}\|_{\mathcal{Q}}^2 = \min!$$

Finally, the new value is computed as

$$q_{n+1} = q_n + h \sum_{j=1}^{s} b_j \dot{q}_{n,j}, \qquad u_{n+1} = \Phi(q_{n+1}).$$

Global error of Runge-Kutta method of order p

Under the same stepsize restriction as for the Euler method,

$$\|u_n - y(t_n)\| = O(\delta + h^p), \qquad 0 \le t_n \le \bar{t},$$

obtained from the local error bound via Lady Windermere's fan with error propagation by the exact flow

Outline

Choice of the regularization parameter ε

For all $\varepsilon > 0$, let $\dot{q}_n(\varepsilon)$ be such that

$$\delta_n(\varepsilon)^2 := \|\Phi'(q_n)\dot{q}_n(\varepsilon) - f(u_n)\|^2 + \varepsilon^2 \|\dot{q}_n(\varepsilon)\|_{\mathcal{Q}}^2 = \text{min!}$$

We know that $\delta_n(\cdot)$ grows monotonically: $\dfrac{d\delta_n^2}{d\varepsilon^2} = \|\dot{q}_n\|_{\mathcal{Q}}^2$

For a tiny $\varepsilon_\star \ll \varepsilon_{n-1}$, compute $\delta_n(\varepsilon_\star)$.

With a given threshold $\delta^{\text{tol}} > 0$, set $\delta_n^{\text{opt}} = \max(17\,\delta_n(\varepsilon_\star), \delta^{\text{tol}})$.

We aim at choosing ε_n such that

$$\delta_n(\varepsilon_n) \approx \delta_n^{\text{opt}}$$

and use 1 or 2 Newton iterations

$$\left(\varepsilon_n^{k+1}\right)^2 = \left(\varepsilon_n^k\right)^2 - \frac{\delta_n(\varepsilon_n^k)^2 - \left(\delta_n^{\text{opt}}\right)^2}{\|\dot{q}_n(\varepsilon_n^k)\|_{\mathcal{Q}}^2}.$$

Choice of the stepsize

We aim at

$$\|\Phi''(q_n)[h_n\dot{q}_n, h_n\dot{q}_n]\| \approx h_n\delta_n$$

which is satisfied for the choice

$$h_n = \frac{h_n^0\delta_n}{\|(\Phi'(q_n + h_n^0\dot{q}_n) - \Phi'(q_n))\dot{q}_n\|}.$$

This is combined with a standard Runge–Kutta stepsize selection.

Compute q_{n+1} and $u_{n+1} = \Phi(q_{n+1})$ by a regularized Runge–Kutta step with the proposed regularization parameter ε_n and the proposed stepsize h_n. The local error estimate from an embedded pair of Runge–Kutta methods should not substantially exceed $h_n\delta_n$; else the step is rejected and repeated with a reduced stepsize.

Outline

Further topics

in our paper, but here not addressed in detail:

- Schrödinger equation (PDE: not Lipschitz)
- Enforcing conserved quantities

Outline

Approximating the flow map of an ODE system

Consider the Lotka-Volterra predator-prey model

$$\dot{x} = x - xy, \qquad \dot{y} \;= xy - y$$

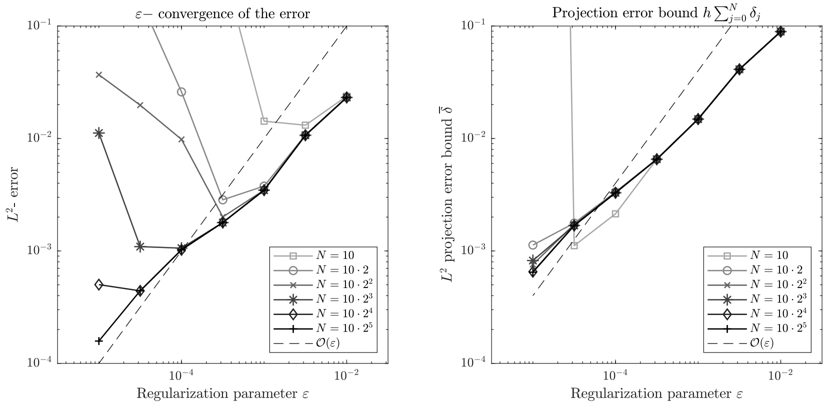
with initial values in $D = [0.5, 2.5]^2$.

The flow map $\varphi_t : D \to \mathbb{R}^2$ at time t is considered as an element of the Hilbert space $\mathcal{H} = L^2(D)^2$. It satisfies the differential equation on $\mathcal{H}$

$$\frac{d}{dt}\varphi_t = f(\varphi_t), \qquad \varphi_0 = \mathrm{Id}.$$

The flow map φ_t is approximated by a neural network with three hidden layers, each with a depth of four neurons. The activation function on each layer is the sigmoid function $\sigma(x) = e^x/(1 + e^x)$.

Numerical experiment



Outline