An introduction to computer-assisted proofs via a posteriori validation

Maxime Breden
CMAP, Ecole polytechnique

MAX team seminar, March 18, 2024

Computer-assisted proofs / a posteriori validation

Computer-assisted proofs / a posteriori validation

- Objective: prove *quantitative* theorems about some *specific* solutions of a given ODE or PDE, using numerical simulations.

Computer-assisted proofs / a posteriori validation

- Objective: prove *quantitative* theorems about some *specific* solutions of a given ODE or PDE, using numerical simulations.
  - steady states
  - periodic orbits
  - eigenvalues/eigenfunctions
  - invariant manifolds
  - connecting orbits
  - traveling waves
  - ...

Computer-assisted proofs / a posteriori validation

- Objective: prove *quantitative* theorems about some *specific* solutions of a given ODE or PDE, using numerical simulations.
    - steady states
    - periodic orbits
    - eigenvalues/eigenfunctions
    - invariant manifolds
    - connecting orbits
    - traveling waves
    - ...
- Starting from a numerical approximation, we prove the existence of an exact solution nearby.

Computer-assisted proofs / a posteriori validation

- Objective: prove *quantitative* theorems about some *specific* solutions of a given ODE or PDE, using numerical simulations.
  - steady states
  - periodic orbits
  - eigenvalues/eigenfunctions
  - invariant manifolds
  - connecting orbits
  - traveling waves
  - ...
- Starting from a numerical approximation, we prove the existence of an exact solution nearby.
- Such computer-assisted approaches use ideas going back to [Lanford '82; Nakao '88; Plum '90; ...].

Computer-assisted proofs / a posteriori validation

- Objective: prove *quantitative* theorems about some *specific* solutions of a given ODE or PDE, using numerical simulations.
  - steady states
  - periodic orbits
  - eigenvalues/eigenfunctions
  - invariant manifolds
  - connecting orbits
  - traveling waves
  - ...
- Starting from a numerical approximation, we prove the existence of an exact solution nearby.
- Such computer-assisted approaches use ideas going back to [Lanford '82; Nakao '88; Plum '90; ...].
- Possible motivation: prove theorems that cannot be proven by "classical" pen-and-paper methods.

Computer-assisted proofs / a posteriori validation

- Objective: prove *quantitative* theorems about some *specific* solutions of a given ODE or PDE, using numerical simulations.
  - steady states
  - periodic orbits
  - eigenvalues/eigenfunctions
  - invariant manifolds
  - connecting orbits
  - traveling waves
  - ...
- Starting from a numerical approximation, we prove the existence of an exact solution nearby.
- Such computer-assisted approaches use ideas going back to [Lanford '82; Nakao '88; Plum '90; ...].
- Possible motivation: prove theorems that cannot be proven by "classical" pen-and-paper methods.
- Alternate viewpoint: these computer-assisted techniques can be seen as a way to guarantee/certify the output of some numerical simulations.

Outline

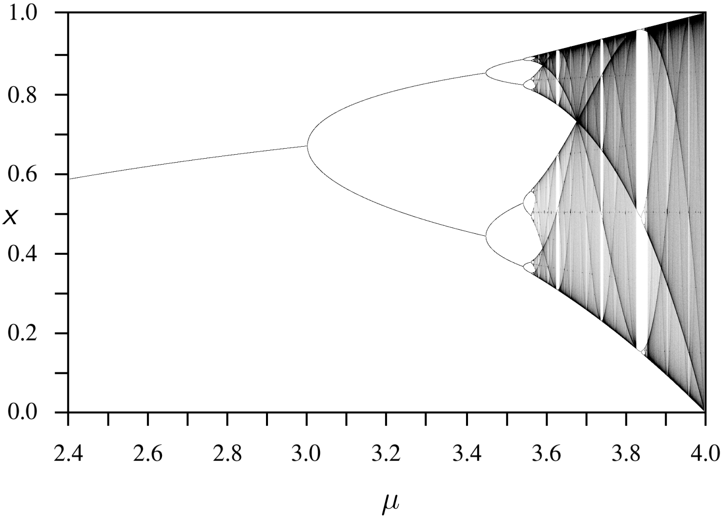
Outline

Motivation: periodic orbits and chaos

Consider the sequence given by the logistic map: $x_{n+1} = \mu x_n(1 - x_n)$.

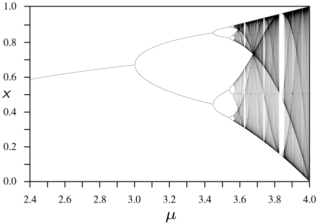
Motivation: periodic orbits and chaos

Consider the sequence given by the logistic map: $x_{n+1} = \mu x_n(1 - x_n)$.



Motivation: periodic orbits and chaos

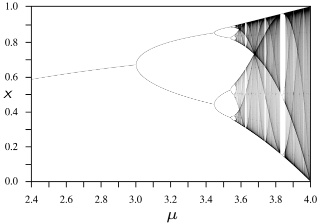
Consider the sequence given by the logistic map: $x_{n+1} = \mu x_n(1 - x_n)$.



- How to characterize the observed dynamics, which becomes more and more complex when μ gets close to 4?

Motivation: periodic orbits and chaos

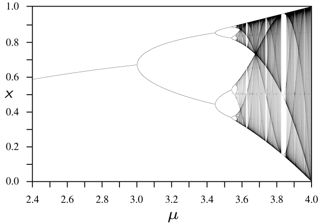
Consider the sequence given by the logistic map: $x_{n+1} = \mu x_n(1 - x_n)$.



- How to characterize the observed dynamics, which becomes more and more complex when μ gets close to 4? Notion of *chaos*.

Motivation: periodic orbits and chaos

Consider the sequence given by the logistic map: $x_{n+1} = \mu x_n(1 - x_n)$.



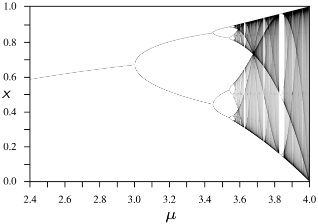
- How to characterize the observed dynamics, which becomes more and more complex when μ gets close to 4? Notion of *chaos*.

Theorem [Sharkovsky '64, Li York '75]

"The existence of a period 3 orbit implies chaos"

Motivation: periodic orbits and chaos

Consider the sequence given by the logistic map: $x_{n+1} = \mu x_n(1 - x_n)$.



- How to characterize the observed dynamics, which becomes more and more complex when μ gets close to 4? Notion of *chaos*.

Theorem [Sharkovsky '64, Li York '75]

"The existence of a period 3 orbit implies chaos"

- For a given value of μ, how can we prove the existence of a period 3 orbit, in order to apply the above theorem?

On the hunt for period 3 orbits

$$x_{n+1} = \mu x_n(1 - x_n)$$

On the hunt for period 3 orbits

$$x_{n+1} = \mu x_n(1 - x_n)$$

- We start by looking numerically for a period 3 orbit.

On the hunt for period 3 orbits

$$x_{n+1} = \mu x_n(1 - x_n)$$

- We start by looking numerically for a period 3 orbit.
- To do so, we can consider the map $F : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$F(x_0, x_1, x_2) = \begin{pmatrix} \mu x_0(1 - x_0) - x_1 \\ \mu x_1(1 - x_1) - x_2 \\ \mu x_2(1 - x_2) - x_0 \end{pmatrix}.$$

  If we manage to find a zero of F (such that $x_0 \neq x_1 \neq x_2$), we then have a period 3 orbit.

On the hunt for period 3 orbits

$$x_{n+1} = \mu x_n(1 - x_n)$$

- We start by looking numerically for a period 3 orbit.
- To do so, we can consider the map $F : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$F(x_0, x_1, x_2) = \begin{pmatrix} \mu x_0(1 - x_0) - x_1 \\ \mu x_1(1 - x_1) - x_2 \\ \mu x_2(1 - x_2) - x_0 \end{pmatrix}.$$

  If we manage to find a zero of F (such that $x_0 \neq x_1 \neq x_2$), we then have a period 3 orbit.
- Numerically, it is easy to find an "approximate solution" $\bar{X} = (\bar{x}_0, \bar{x}_1, \bar{x}_2)$ such that $F(\bar{X}) \approx 0$.

On the hunt for period 3 orbits

$$x_{n+1} = \mu x_n(1 - x_n)$$

- We start by looking numerically for a period 3 orbit.
- To do so, we can consider the map $F : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$F(x_0, x_1, x_2) = \begin{pmatrix} \mu x_0(1 - x_0) - x_1 \\ \mu x_1(1 - x_1) - x_2 \\ \mu x_2(1 - x_2) - x_0 \end{pmatrix}.$$

  If we manage to find a zero of F (such that $x_0 \neq x_1 \neq x_2$), we then have a period 3 orbit.
- Numerically, it is easy to find an "approximate solution" $\bar{X} = (\bar{x}_0, \bar{x}_1, \bar{x}_2)$ such that $F(\bar{X}) \approx 0$.
- How to rigorously prove the existence of this zero of F?

We need proof!

$$F : \mathbb{R}^3 \to \mathbb{R}^3 \qquad F(\bar{X}) \approx 0.$$

We need proof!

$$F : \mathbb{R}^3 \to \mathbb{R}^3 \qquad F(\bar{X}) \approx 0.$$

- We want to prove **a posteriori** the existence of a zero of F close to $\bar{X}$.

We need proof!

$$F : \mathbb{R}^3 \to \mathbb{R}^3 \qquad F(\bar{X}) \approx 0.$$

- We want to prove **a posteriori** the existence of a zero of F close to $\bar{X}$.

Theorem *(à la Newton-Kantorovich)*

Let $\varepsilon, K, L > 0$ such that

$$\|F(\bar{X})\| \leq \varepsilon$$
$$\|DF(\bar{X})^{-1}\| \leq \kappa$$
$$\|DF(X) - DF(\bar{X})\| \leq L\|X - \bar{X}\| \qquad \forall\, X \in \mathbb{R}^3.$$

If

$$\varepsilon < \frac{1}{2\kappa^2 L},$$

then F has a unique zero X^* satisfying $\|X^* - \bar{X}\| \leq r$, $r = \frac{1-\sqrt{1-2\kappa^2 L\varepsilon}}{\kappa L}$.

We need proof!

$$F : \mathbb{R}^3 \to \mathbb{R}^3 \qquad F(\bar{X}) \approx 0.$$

- We want to prove **a posteriori** the existence of a zero of F close to $\bar{X}$.

Theorem *(à la Newton-Kantorovich)*

Let $\varepsilon, K, L > 0$ such that

$$\|F(\bar{X})\| \leq \varepsilon$$
$$\|DF(\bar{X})^{-1}\| \leq \kappa$$
$$\|DF(X) - DF(\bar{X})\| \leq L\|X - \bar{X}\| \qquad \forall\, X \in \mathbb{R}^3.$$

If

$$\varepsilon < \frac{1}{2\kappa^2 L},$$

then F has a unique zero X^* satisfying $\|X^* - \bar{X}\| \leq r$, $r = \frac{1-\sqrt{1-2\kappa^2 L\varepsilon}}{\kappa L}$.

Proof : $T : X \mapsto X - DF(\bar{X})^{-1}F(X)$ is a contraction on the closed ball of center $\bar{X}$ and radius r.

A frightening example

- Can we really trust floating-point arithmetic?

A frightening example

- Can we really trust floating-point arithmetic?
- Consider the following expression [Rump '94]

$$g(a,b) = 333.75b^6 + a^2(11a^2b^2 - b^6 - 121b^4 - 2) + 5.5b^8 + \frac{a}{2b},$$

evaluated for $a = 77617$ and $b = 33096$, with various precisions.

A frightening example

- Can we really trust floating-point arithmetic?
- Consider the following expression [Rump '94]

$$g(a,b) = 333.75b^6 + a^2(11a^2b^2 - b^6 - 121b^4 - 2) + 5.5b^8 + \frac{a}{2b},$$

evaluated for $a = 77617$ and $b = 33096$, with various precisions.

- We have to be wary of round-off errors, especially if we claim to have proven a theorem based on some numerical computations!

A frightening example

- Can we really trust floating-point arithmetic?

- Consider the following expression [Rump '94]

$$g(a, b) = 333.75b^6 + a^2(11a^2b^2 - b^6 - 121b^4 - 2) + 5.5b^8 + \frac{a}{2b},$$

evaluated for $a = 77617$ and $b = 33096$, with various precisions.

- We have to be wary of round-off errors, especially if we claim to have proven a theorem based on some numerical computations!

- In our "proof" of existence of a period 3 orbit, how can we be certain that the quantity ε that we numerically evaluated really bounds $\|F(\bar{X})\|$, or that $\varepsilon < \frac{1}{2\kappa^2 L}$?

Calling interval arithmetic to the rescue

Calling interval arithmetic to the rescue

- Let $\mathbb{F}$ bet a set of floating point numbers, corresponding to the (finite!) set of real numbers that the computer can represent with a given precision, and

$$\triangledown, \triangle : \mathbb{R} \to \mathbb{F},$$

the round-down and round-up operators.

Calling interval arithmetic to the rescue

- Let $\mathbb{F}$ bet a set of floating point numbers, corresponding to the (finite!) set of real numbers that the computer can represent with a given precision, and
$$\triangledown, \triangle : \mathbb{R} \to \mathbb{F},$$
the round-down and round-up operators.
- Example: consider $x = 0.1$. In base 2, x writes
$$x = (1.10011001100110011001100...)_2 \times 2^{-4}.$$
With 8 bits of precision (for the mantissa), we have
$$\triangledown(x) = (1.1001100)_2 \times 2^{-4} \quad \text{and} \quad \triangle(x) = (1.1001101)_2 \times 2^{-4}.$$

Calling interval arithmetic to the rescue

- Let $\mathbb{F}$ bet a set of floating point numbers, corresponding to the (finite!) set of real numbers that the computer can represent with a given precision, and

$$\triangledown, \triangle : \mathbb{R} \to \mathbb{F},$$

the round-down and round-up operators.

- Instead of using floats, we now represent each real number by an interval which contains it:

$$x \in \mathbb{R} \quad \to \quad [x] := [\triangledown(x)\,,\, \triangle(x)]\,.$$

Calling interval arithmetic to the rescue

- Let $\mathbb{F}$ bet a set of floating point numbers, corresponding to the (finite!) set of real numbers that the computer can represent with a given precision, and

$$\triangledown, \triangle : \mathbb{R} \to \mathbb{F},$$

the round-down and round-up operators.

- Instead of using floats, we now represent each real number by an interval which contains it:

$$x \in \mathbb{R} \quad \to \quad [x] := [\triangledown(x)\,,\, \triangle(x)]\,.$$

- On can then extend the elementary operations $(+, -, \times, \div)$ to intervals, in such a way that the result always contain the true value:

$$x + y \quad \to \quad [x]\ [+]\ [y],$$

where $[+]$ is defined as follows

$$[x]\ [+]\ [y] := [\triangledown\left(\triangledown(x) + \triangledown(y)\right)\,,\, \triangle\left(\triangle(x) + \triangle(y)\right)]\,.$$

Calling interval arithmetic to the rescue

- Let $\mathbb{F}$ bet a set of floating point numbers, corresponding to the (finite!) set of real numbers that the computer can represent with a given precision, and
$$\triangledown, \triangle : \mathbb{R} \to \mathbb{F},$$
the round-down and round-up operators.
- Instead of using floats, we now represent each real number by an interval which contains it:
$$x \in \mathbb{R} \quad \to \quad [x] := [\triangledown(x)\,,\, \triangle(x)]\,.$$
- On can then extend the elementary operations $(+, -, \times, \div)$ to intervals, in such a way that the result always contain the true value:
$$x + y \quad \to \quad [x]\ [+]\ [y],$$
where $[+]$ is defined as follows (doable in practice, IEEE 754 standard)
$$[x]\ [+]\ [y] := [\triangledown\left(\triangledown(x) + \triangledown(y)\right)\,,\, \triangle\left(\triangle(x) + \triangle(y)\right)]\,.$$

Calling interval arithmetic to the rescue

- Let $\mathbb{F}$ bet a set of floating point numbers, corresponding to the (finite!) set of real numbers that the computer can represent with a given precision, and
$$\nabla, \triangle : \mathbb{R} \to \mathbb{F},$$
the round-down and round-up operators.
- Instead of using floats, we now represent each real number by an interval which contains it:
$$x \in \mathbb{R} \quad \to \quad [x] := [\nabla(x)\,,\, \triangle(x)]\,.$$
- On can then extend the elementary operations $(+, -, \times, \div)$ to intervals, in such a way that the result always contain the true value:
$$x + y \quad \to \quad [x]\ [+]\ [y],$$
where $[+]$ is defined as follows (doable in practice, IEEE 754 standard)
$$[x]\ [+]\ [y] := [\nabla\left(\nabla(x) + \nabla(y)\right)\,,\, \triangle\left(\triangle(x) + \triangle(y)\right)]\,.$$
- We then have $\boldsymbol{x} + \boldsymbol{y} \in [\boldsymbol{x}]\ [+]\ [\boldsymbol{y}]$.

A computer-assisted proof a chaos, summary

$$x_{n+1} = \mu x_n(1 - x_n)$$

A computer-assisted proof a chaos, summary

$$x_{n+1} = \mu x_n(1 - x_n)$$

1. We reformulate the search of a period 3 orbit as a zero-finding problem $\boldsymbol{F}(\boldsymbol{X}) = \mathbf{0}$.

A computer-assisted proof a chaos, summary

$$x_{n+1} = \mu x_n(1 - x_n)$$

1. We reformulate the search of a period 3 orbit as a zero-finding problem $\boldsymbol{F}(\boldsymbol{X}) = \mathbf{0}$.

2. We numerically find an **approximate solution**.

A computer-assisted proof a chaos, summary

$$x_{n+1} = \mu x_n(1 - x_n)$$

1. We reformulate the search of a period 3 orbit as a zero-finding problem $\boldsymbol{F}(\boldsymbol{X}) = \mathbf{0}$.

2. We numerically find an **approximate solution**.

3. We **estimate a posteriori**

$$\|F(\bar{X})\|, \quad \|DF(\bar{X})^{-1}\| \quad \text{and} \quad \|D^2F(X)\|,$$

and do so rigorously using **interval arithmetic**.

A computer-assisted proof a chaos, summary

$$x_{n+1} = \mu x_n(1 - x_n)$$

1. We reformulate the search of a period 3 orbit as a zero-finding problem $\boldsymbol{F}(\boldsymbol{X}) = \mathbf{0}$.

2. We numerically find an **approximate solution**.

3. We **estimate a posteriori**

$$\|F(\bar{X})\|, \quad \|DF(\bar{X})^{-1}\| \quad \text{and} \quad \|D^2F(X)\|,$$

and do so rigorously using **interval arithmetic**.

4. We use these estimates to prove that

$$\boldsymbol{T} : \boldsymbol{X} \mapsto \boldsymbol{X} - \boldsymbol{DF}(\bar{\boldsymbol{X}})^{-1}\boldsymbol{F}(\boldsymbol{X})$$

is a contraction on a small neighborhood of $\bar{X}$.

Outline

How to use these ideas in a broader context?

How to use these ideas in a broader context?

1. Reformulate the problem we are interested in (ODE, PDE, etc) in the form $F(X) = 0$.
   - Several possible choices for F.
   - We also need to chose a Banach space $\mathcal{X}$, and in particular a norm.

How to use these ideas in a broader context?

1. Reformulate the problem we are interested in (ODE, PDE, etc) in the form $F(X) = 0$.
   - Several possible choices for F.
   - We also need to chose a Banach space $\mathcal{X}$, and in particular a norm.

2. Find numerically an approximate zero $\bar{X}$.
   - Choice of discretization method, of a finite dimensional space $\mathcal{X}_h$ in which we look for the approximate solution.

How to use these ideas in a broader context?

1. Reformulate the problem we are interested in (ODE, PDE, etc) in the form $F(X) = 0$.
   - Several possible choices for F.
   - We also need to chose a Banach space $\mathcal{X}$, and in particular a norm.

2. Find numerically an approximate zero $\bar{X}$.
   - Choice of discretization method, of a finite dimensional space $\mathcal{X}_h$ in which we look for the approximate solution.

3. Estimate a posteriori

$$\|F(\bar{X})\|, \quad \|DF(\bar{X})^{-1}\| \quad \text{and} \quad \|D^2F(X)\|.$$

How to use these ideas in a broader context?

1. Reformulate the problem we are interested in (ODE, PDE, etc) in the form $F(X) = 0$.
   - Several possible choices for F.
   - We also need to chose a Banach space $\mathcal{X}$, and in particular a norm.

2. Find numerically an approximate zero $\bar{X}$.
   - Choice of discretization method, of a finite dimensional space $\mathcal{X}_h$ in which we look for the approximate solution.

3. Estimate a posteriori

$$\|F(\bar{X})\|, \quad \|DF(\bar{X})^{-1}\| \quad \text{and} \quad \|D^2F(X)\|.$$

   - **The main difficulty lies in controlling $\|DF(\bar{X})^{-1}\|$.**

A new validation criteria

Theorem *à la Newton-Kantorovich* bis

Let $\varepsilon, \kappa, L, \delta > 0$ such that

$$\|F(\bar{X})\| \leq \varepsilon, \quad \|A\| \leq \kappa, \quad \|DF(X) - DF(\bar{X})\| \leq L\|X - \bar{X}\|,$$

$$\|I - ADF(\bar{X})\| \leq \delta < 1.$$

If

$$\varepsilon < \frac{(1-\delta)^2}{2\kappa^2 L},$$

then F has a unique zero X^* satisfying $\|X^* - \bar{X}\| \leq r$, $r = \frac{1-\delta-\sqrt{(1-\delta)^2-2\kappa^2 L}}{\kappa L}$.

A new validation criteria

Theorem *à la Newton-Kantorovich* bis

Let $\varepsilon, \kappa, L, \delta > 0$ such that

$$\|F(\bar{X})\| \leq \varepsilon, \quad \|A\| \leq \kappa, \quad \|DF(X) - DF(\bar{X})\| \leq L\|X - \bar{X}\|,$$

$$\|I - ADF(\bar{X})\| \leq \delta < 1.$$

If

$$\varepsilon < \frac{(1-\delta)^2}{2\kappa^2 L},$$

then F has a unique zero X^* satisfying $\|X^* - \bar{X}\| \leq r$, $r = \frac{1-\delta-\sqrt{(1-\delta)^2-2\kappa^2 L}}{\kappa L}$.

- An equivalent way to interpret this strategy is to say that we replace the former fixed-point operator $T : x \mapsto x - DF(\bar{x})^{-1}F(x)$ by

$$\tilde{T} : x \mapsto x - AF(x).$$

Setting for validated integration of ODEs

$$\begin{cases} u'(t) = f(u(t)) & t \in [0, 2\tau] \\ u(0) = u^{in} \end{cases}$$

with $f : \mathbb{R}^d \to \mathbb{R}^d$ smooth and $\tau > 0$ fixed.

Setting for validated integration of ODEs

$$\begin{cases} u'(t) = f(u(t)) & t \in [0, 2\tau] \\ u(0) = u^{in} \end{cases}$$

with $f : \mathbb{R}^d \to \mathbb{R}^d$ smooth and $\tau > 0$ fixed.

- Goal: given an approximate solution $\bar{u} : [0, 2\tau] \to \mathbb{R}^d$, prove that the exact solution u satisfies $\|u - \bar{u}\| \leq r$ for some explicit r.

Setting for validated integration of ODEs

$$\begin{cases} u'(t) = f(u(t)) & t \in [0, 2\tau] \\ u(0) = u^{in} \end{cases}$$

with $f : \mathbb{R}^d \to \mathbb{R}^d$ smooth and $\tau > 0$ fixed.

- Goal: given an approximate solution $\bar{u} : [0, 2\tau] \to \mathbb{R}^d$, prove that the exact solution u satisfies $\|u - \bar{u}\| \leq r$ for some explicit r.
- Main idea for the zero-finding problem:

$$F(u)(t) = u(t) - \left(u^{in} + \int_0^t f(u(s)) \mathrm{d}s \right).$$

Setting for validated integration of ODEs

$$\begin{cases} u'(t) = f(u(t)) & t \in [0, 2\tau] \\ u(0) = u^{in} \end{cases}$$

with $f : \mathbb{R}^d \to \mathbb{R}^d$ smooth and $\tau > 0$ fixed.

- Goal: given an approximate solution $\bar{u} : [0, 2\tau] \to \mathbb{R}^d$, prove that the exact solution u satisfies $\|u - \bar{u}\| \leq r$ for some explicit r.
- Main idea for the zero-finding problem:

$$F(u)(t) = u(t) - \left(u^{in} + \int_0^t f(u(s)) \mathrm{d}s \right).$$

- Key observation:

$$DF(\bar{u})(h)(t) = h(t) - \int_0^t Df(\bar{u}(s))h(s) \mathrm{d}s,$$

i.e., $DF(\bar{u})$ is a compact perturbation of the identity.

Chebyshev series

$$\begin{cases} u'(t) = \tau f(u(t)) & t \in [-1,1] \\ u(-1) = u^{in} \end{cases}$$

Chebyshev series

$$\begin{cases} u'(t) = \tau f(u(t)) & t \in [-1,1] \\ u(-1) = u^{in} \end{cases}$$

- Look for the solution as a Chebyshev series:

$$u(t) = u_0 + 2\sum_{n=1}^{\infty} u_n T_n(t), \qquad T_n(\cos\theta) = \cos(n\theta).$$

Chebyshev series

$$\begin{cases} u'(t) = \tau f(u(t)) & t \in [-1,1] \\ u(-1) = u^{in} \end{cases}$$

- Look for the solution as a Chebyshev series:

$$u(t) = u_0 + 2\sum_{n=1}^{\infty} u_n T_n(t), \qquad T_n(\cos\theta) = \cos(n\theta).$$

- The unknown is the sequence $\boldsymbol{u} = (u_n)_{n\geq 0}$ of Chebyshev coefficients.

Chebyshev series

$$\begin{cases} u'(t) = \tau f(u(t)) & t \in [-1,1] \\ u(-1) = u^{in} \end{cases}$$

- Look for the solution as a Chebyshev series:

$$u(t) = u_0 + 2\sum_{n=1}^{\infty} u_n T_n(t), \qquad T_n(\cos\theta) = \cos(n\theta).$$

- The unknown is the sequence $\boldsymbol{u} = (u_n)_{n\geq 0}$ of Chebyshev coefficients.
- By plugging the Chebyshev series ansatz into

$$u(t) - \left(u^{in} + \tau \int_{-1}^{t} f(u(s))\mathrm{d}s\right) = 0,$$

we obtain our $F(\boldsymbol{u}) = 0$ problem.

Chebyshev series

$$\begin{cases} u'(t) = \tau f(u(t)) & t \in [-1,1] \\ u(-1) = u^{in} \end{cases}$$

- Look for the solution as a Chebyshev series:

$$u(t) = u_0 + 2\sum_{n=1}^{\infty} u_n T_n(t), \qquad T_n(\cos\theta) = \cos(n\theta).$$

- The unknown is the sequence $\boldsymbol{u} = (u_n)_{n\geq 0}$ of Chebyshev coefficients.
- By plugging the Chebyshev series ansatz into

$$u(t) - \left(u^{in} + \tau \int_{-1}^{t} f(u(s))\mathrm{d}s\right) = 0,$$

  we obtain our $F(\boldsymbol{u}) = 0$ problem.
- The approximate solution $\bar{\boldsymbol{u}}$ is taken as a truncated Chebyshev series.

Chebyshev series

$$\begin{cases} u'(t) = \tau f(u(t)) & t \in [-1,1] \\ u(-1) = u^{in} \end{cases}$$

- Look for the solution as a Chebyshev series:

$$u(t) = u_0 + 2\sum_{n=1}^{\infty} u_n T_n(t), \qquad T_n(\cos\theta) = \cos(n\theta).$$

- The unknown is the sequence $\boldsymbol{u} = (u_n)_{n\geq 0}$ of Chebyshev coefficients.
- By plugging the Chebyshev series ansatz into

$$u(t) - \left(u^{in} + \tau \int_{-1}^{t} f(u(s)) \mathrm{d}s \right) = 0,$$

  we obtain our $F(\boldsymbol{u}) = 0$ problem.
- The approximate solution $\bar{\boldsymbol{u}}$ is taken as a truncated Chebyshev series.
- We look for the exact solution in the space $\ell^1_\nu := \{\boldsymbol{u}, \left\|\boldsymbol{u}\right\|_\nu < \infty\}$,

$$\left\|\boldsymbol{u}\right\|_\nu := |u_0| + 2\sum_{n=1}^{\infty} |u_n|\, \nu^n, \qquad \nu \geq 1.$$

Why Chebyshev series?

$$u(t) - \left(u^{in} + \tau \int_{-1}^{t} f(u(s)) \mathrm{d}s \right) = 0.$$

Why Chebyshev series?

$$u(t) - \left(u^{in} + \tau \int_{-1}^{t} f(u(s))\mathrm{d}s \right) = 0.$$

- Excellent approximation properties (similar to Fourier series for periodic functions).

Why Chebyshev series?

$$u(t) - \left(u^{in} + \tau \int_{-1}^{t} f(u(s))\mathrm{d}s \right) = 0.$$

- Excellent approximation properties (similar to Fourier series for periodic functions).
- Easy formulation of the antiderivative allowing to "see" the compactness

$$\int T_n = \frac{1}{2} \left(\frac{1}{n+1} T_{n+1} - \frac{1}{n-1} T_{n-1} \right).$$

Why Chebyshev series?

$$u(t) - \left(u^{in} + \tau \int_{-1}^{t} f(u(s))\mathrm{d}s\right) = 0.$$

- Excellent approximation properties (similar to Fourier series for periodic functions).
- Easy formulation of the antiderivative allowing to "see" the compactness

$$\int T_n = \frac{1}{2}\left(\frac{1}{n+1}T_{n+1} - \frac{1}{n-1}T_{n-1}\right).$$

- Efficient computations of nonlinearities using the FFT.

Why Chebyshev series?

$$u(t) - \left(u^{in} + \tau \int_{-1}^{t} f(u(s))\mathrm{d}s \right) = 0.$$

- Excellent approximation properties (similar to Fourier series for periodic functions).
- Easy formulation of the antiderivative allowing to "see" the compactness

$$\int T_n = \frac{1}{2}\left(\frac{1}{n+1} T_{n+1} - \frac{1}{n-1} T_{n-1} \right).$$

- Efficient computations of nonlinearities using the FFT.
  - Computing $\left\| F(\bar{\boldsymbol{u}}) \right\|_{\nu}$ is rather straightforward.

Why Chebyshev series?

$$u(t) - \left(u^{in} + \tau \int_{-1}^{t} f(u(s))\mathrm{d}s\right) = 0.$$

- Excellent approximation properties (similar to Fourier series for periodic functions).
- Easy formulation of the antiderivative allowing to "see" the compactness

$$\int T_n = \frac{1}{2}\left(\frac{1}{n+1}T_{n+1} - \frac{1}{n-1}T_{n-1}\right).$$

- Efficient computations of nonlinearities using the FFT.
  - Computing $\left\|F(\bar{\boldsymbol{u}})\right\|_\nu$ is rather straightforward.
- ℓ^1_ν is a Banach algebra: $\left\|\boldsymbol{u} * \boldsymbol{v}\right\|_\nu \leq \left\|\boldsymbol{u}\right\|_\nu \left\|\boldsymbol{v}\right\|_\nu$.

Why Chebyshev series?

$$u(t) - \left(u^{in} + \tau \int_{-1}^{t} f(u(s))\mathrm{d}s \right) = 0.$$

- Excellent approximation properties (similar to Fourier series for periodic functions).
- Easy formulation of the antiderivative allowing to "see" the compactness

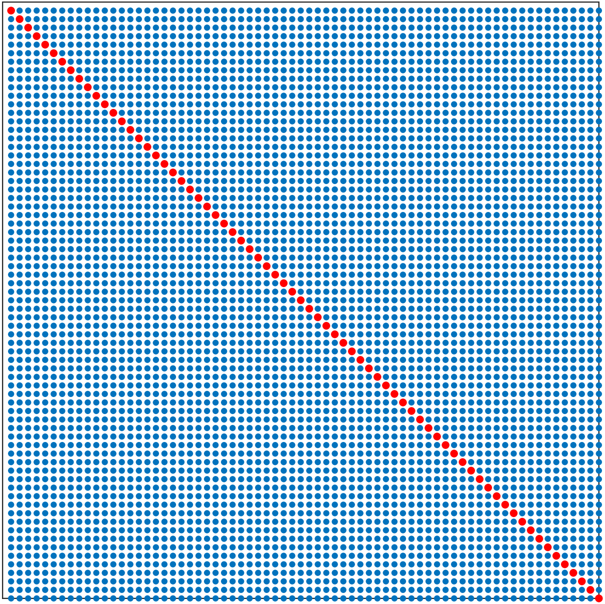
$$\int T_n = \frac{1}{2}\left(\frac{1}{n+1} T_{n+1} - \frac{1}{n-1} T_{n-1} \right).$$

- Efficient computations of nonlinearities using the FFT.
  - Computing $\left\| F(\bar{\boldsymbol{u}}) \right\|_\nu$ is rather straightforward.
- ℓ^1_ν is a Banach algebra: $\left\| \boldsymbol{u} * \boldsymbol{v} \right\|_\nu \leq \left\| \boldsymbol{u} \right\|_\nu \left\| \boldsymbol{v} \right\|_\nu$.
  - Simplifies the estimation of $\left\| D^2 F(\boldsymbol{u}) \right\|_\nu$ for $\boldsymbol{u}$ in a neighborhood of $\bar{\boldsymbol{u}}$.

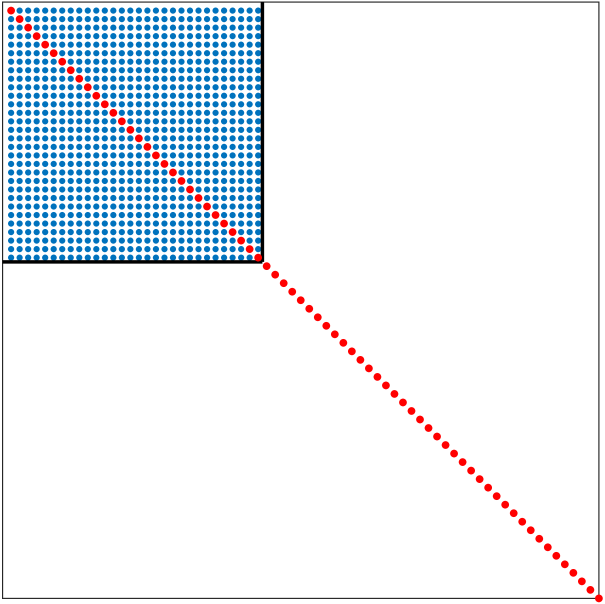
How to construct the approximate inverse A

How to construct the approximate inverse A

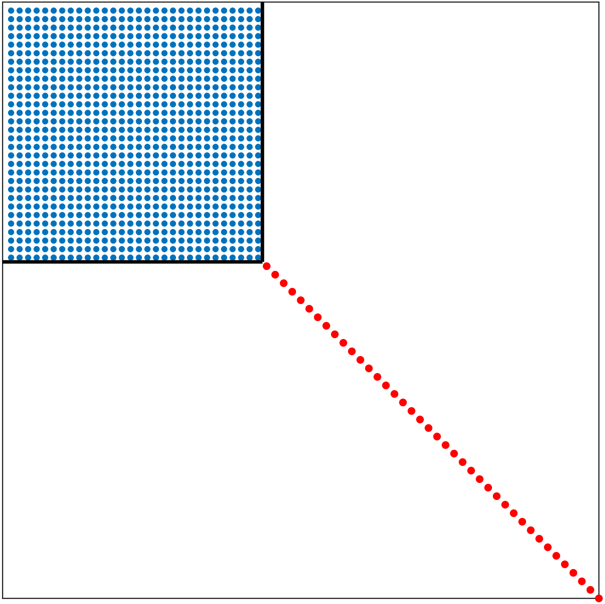
$$DF(\bar{\boldsymbol{u}}) =$$



How to construct the approximate inverse A

$DF(\bar{\boldsymbol{u}}) \approx$ 

How to construct the approximate inverse A

$A :=$ 

Quality of this approximate inverse

Quality of this approximate inverse

- Using this constructing, when keeping the first N Chebyshev modes in the finite block, we get

$$\left\| I - A\, DF(\bar{\boldsymbol{u}}) \right\|_{\nu} \approx \frac{\tau \left\| f'(\bar{\boldsymbol{u}}) \right\|_{\nu}}{N}.$$

Quality of this approximate inverse

- Using this constructing, when keeping the first N Chebyshev modes in the finite block, we get

$$\left\| I - A\, DF(\bar{\boldsymbol{u}}) \right\|_\nu \approx \frac{\tau \left\| f'(\bar{\boldsymbol{u}}) \right\|_\nu}{N}.$$

- Up to taking N large enough, we can therefore get $\left\| I - A\, DF(\bar{\boldsymbol{u}}) \right\|_\nu < 1$, and hope to apply the entire *a posteriori validation* procedure.

Quality of this approximate inverse

- Using this constructing, when keeping the first N Chebyshev modes in the finite block, we get

$$\left\| I - A\, DF(\bar{\boldsymbol{u}}) \right\|_{\nu} \approx \frac{\tau \left\| f'(\bar{\boldsymbol{u}}) \right\|_{\nu}}{N}.$$

- Up to taking N large enough, we can therefore get $\left\| I - A\, DF(\bar{\boldsymbol{u}}) \right\|_{\nu} < 1$, and hope to apply the entire *a posteriori validation* procedure.

- [Lessard Reinhardt '14]

Domain decomposition

Domain decomposition

- It can be helpful to split the solution into several "Chebyshev pieces", by decomposing the time interval: $0 = \tau_0 < \tau_1 < \ldots < \tau_M = \tau$.

Domain decomposition

- It can be helpful to split the solution into several "Chebyshev pieces", by decomposing the time interval: $0 = \tau_0 < \tau_1 < \ldots < \tau_M = \tau$.
- We then look for $u = (u^{(1)}, u^{(2)}, \ldots, u^{(M)})$ so that each $u^{(m)}$ solves the equation on $[\tau_{m-1}, \tau_m]$:

Domain decomposition

- It can be helpful to split the solution into several "Chebyshev pieces", by decomposing the time interval: $0 = \tau_0 < \tau_1 < \ldots < \tau_M = \tau$.
- We then look for $u = (u^{(1)}, u^{(2)}, \ldots, u^{(M)})$ so that each $u^{(m)}$ solves the equation on $[\tau_{m-1}, \tau_m]$:

$$u^{(1)}(t) - \left(u^{in} + \int_0^t f(u^{(1)}(s))\mathrm{d}s \right) = 0 \qquad t \in [0, \tau_1],$$

$$u^{(2)}(t) - \left(u^{(1)}(\tau_1) + \int_{\tau_1}^t f(u^{(2)}(s))\mathrm{d}s \right) = 0 \qquad t \in [\tau_1, \tau_2],$$

$$\vdots$$

$$u^{(M)}(t) - \left(u^{(M-1)}(\tau_{M-1}) + \int_{\tau_{M-1}}^t f(u^{(M)}(s))\mathrm{d}s \right) = 0 \qquad t \in [\tau_{M-1}, \tau].$$

Domain decomposition

- It can be helpful to split the solution into several "Chebyshev pieces", by decomposing the time interval: $0 = \tau_0 < \tau_1 < \ldots < \tau_M = \tau$.
- We then look for $u = (u^{(1)}, u^{(2)}, \ldots, u^{(M)})$ so that each $u^{(m)}$ solves the equation on $[\tau_{m-1}, \tau_m]$:

$$
\begin{aligned}
u^{(1)}(t) - \left(u^{in} + \int_0^t f(u^{(1)}(s))\mathrm{d}s \right) &= 0 \qquad t \in [0, \tau_1], \\
u^{(2)}(t) - \left(u^{(1)}(\tau_1) + \int_{\tau_1}^t f(u^{(2)}(s))\mathrm{d}s \right) &= 0 \qquad t \in [\tau_1, \tau_2], \\
&\vdots \\
u^{(M)}(t) - \left(u^{(M-1)}(\tau_{M-1}) + \int_{\tau_{M-1}}^t f(u^{(M)}(s))\mathrm{d}s \right) &= 0 \qquad t \in [\tau_{M-1}, \tau].
\end{aligned}
$$

- Each $u^{(m)}$ is then represented by a Chebyshev series, and this leads to a big $F(\boldsymbol{u}) = 0$ problem.

Domain decomposition

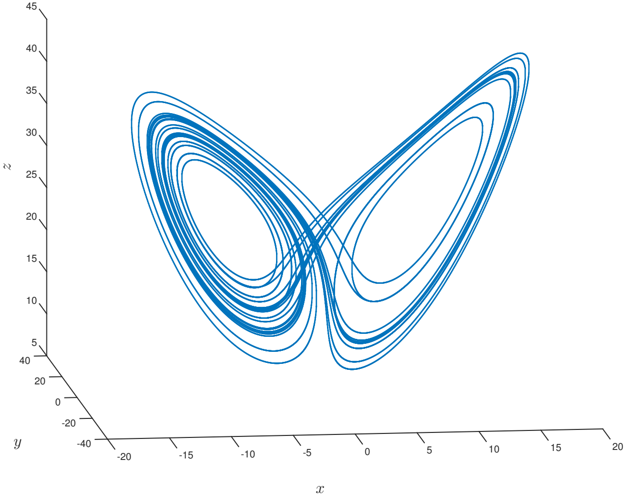
- It can be helpful to split the solution into several "Chebyshev pieces", by decomposing the time interval: $0 = \tau_0 < \tau_1 < \ldots < \tau_M = \tau$.
- We then look for $u = (u^{(1)}, u^{(2)}, \ldots, u^{(M)})$ so that each $u^{(m)}$ solves the equation on $[\tau_{m-1}, \tau_m]$:

$$\begin{aligned}
u^{(1)}(t) - \left(u^{in} + \int_0^t f(u^{(1)}(s))\mathrm{d}s\right) &= 0 \qquad t \in [0, \tau_1], \\
u^{(2)}(t) - \left(u^{(1)}(\tau_1) + \int_{\tau_1}^t f(u^{(2)}(s))\mathrm{d}s\right) &= 0 \qquad t \in [\tau_1, \tau_2], \\
&\vdots \\
u^{(M)}(t) - \left(u^{(M-1)}(\tau_{M-1}) + \int_{\tau_{M-1}}^t f(u^{(M)}(s))\mathrm{d}s\right) &= 0 \qquad t \in [\tau_{M-1}, \tau].
\end{aligned}$$

- Each $u^{(m)}$ is then represented by a Chebyshev series, and this leads to a big $F(\boldsymbol{u}) = 0$ problem.
- [van den Berg Sheombarsing '21]

Some examples from [van den Berg Sheombarsing '21]

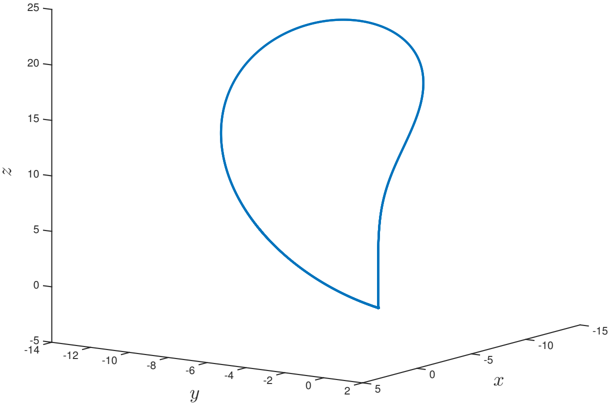
$$x' = 10(x - y)$$
$$y' = 28x - y - xz$$
$$z' = -8z/3 + xy$$



Integration time $\tau \approx 25$

Some examples from [van den Berg Sheombarsing '21]

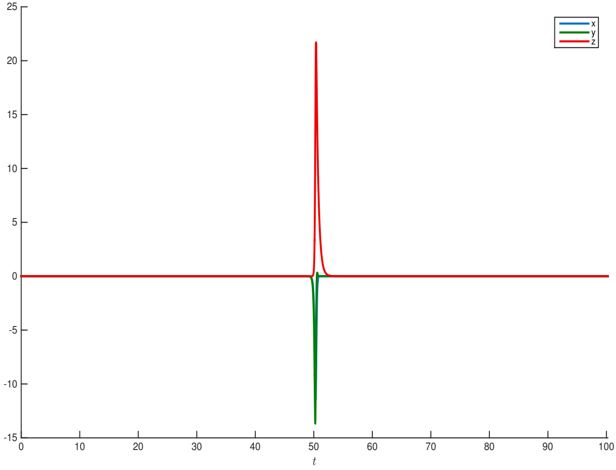
$$x' = 10(x - y)$$
$$y' = 28x - y - xz$$
$$z' = -8z/3 + xy$$



Integration time $\tau \approx 100$

Some examples from [van den Berg Sheombarsing '21]

$$x' = 10(x - y)$$
$$y' = 28x - y - xz$$
$$z' = -8z/3 + xy$$



Integration time $\tau \approx 100$

Some related works

Some related works

- Chebyshev methods for linear ODEs, with special emphasis on studying and potentially reducing computational complexity [Benoit Joldes Mezzarobba '17; Brehard Brisebarre Joldes '18; Brehard '21].

Some related works

- Chebyshev methods for linear ODEs, with special emphasis on studying and potentially reducing computational complexity [Benoit Joldes Mezzarobba '17; Brehard Brisebarre Joldes '18; Brehard '21].

- Many other methods, some of which are more in the spirit of *traditional* numerical methods for ODEs. A particularly successful one is the CAPD::DynSys library [Kapela Mrozek Wilczak Zgliczynski '21].

Outline

A different fixed point reformulation

$$\begin{cases} u'(t) = f(u(t)) & t \in [0, \tau] \\ u(0) = u^{in} \end{cases}$$

A different fixed point reformulation

$$\begin{cases} u'(t) = f(u(t)) & t \in [0, \tau] \\ u(0) = u^{in} \end{cases}$$

- We started by converting the equation into an $F(u) = 0$ problem:

$$F(u)(t) = u(t) - \left(u^{in} + \int_0^t f(u(s))\mathrm{d}s \right),$$

and then into a fixed point problem $T(u) = u - AF(u)$.

A different fixed point reformulation

$$\begin{cases} u'(t) = f(u(t)) & t \in [0, \tau] \\ u(0) = u^{in} \end{cases}$$

- We started by converting the equation into an $F(u) = 0$ problem:

$$F(u)(t) = u(t) - \left(u^{in} + \int_0^t f(u(s))\mathrm{d}s \right),$$

and then into a fixed point problem $T(u) = u - AF(u)$.

- One could also directly get a fixed point problem:

$$\tilde{T}(u)(t) = u^{in} + \int_0^t f(u(s))\mathrm{d}s.$$

A different fixed point reformulation

$$\begin{cases} u'(t) = f(u(t)) & t \in [0,\tau] \\ u(0) = u^{in} \end{cases}$$

- We started by converting the equation into an $F(u) = 0$ problem:

$$F(u)(t) = u(t) - \left(u^{in} + \int_0^t f(u(s))\mathrm{d}s \right),$$

  and then into a fixed point problem $T(u) = u - AF(u)$.

- One could also directly get a fixed point problem:

$$\tilde{T}(u)(t) = u^{in} + \int_0^t f(u(s))\mathrm{d}s.$$

- $\tilde{T}$ has no reason to be contracting near $\bar{u}$, except for τ small.

A different fixed point reformulation

$$\begin{cases} u'(t) - Lu(t) = f(u(t)) - Lu(t) \qquad t \in [0, \tau] \\ u(0) = u^{in} \end{cases}$$

A different fixed point reformulation

$$\begin{cases} u'(t) - Lu(t) = f(u(t)) - Lu(t) & t \in [0, \tau] \\ u(0) = u^{in} \end{cases}$$

- Using Duhamel's principle/the variation of constants formula, we get

$$\tilde{T}(u)(t) = e^{tL}u^{in} + \int_0^t e^{(t-s)L}\left(f(u(s)) - Lu(s)\right) \mathrm{d}s.$$

A different fixed point reformulation

$$\begin{cases} u'(t) - Lu(t) = f(u(t)) - Lu(t) & t \in [0, \tau] \\ u(0) = u^{in} \end{cases}$$

- Using Duhamel's principle/the variation of constants formula, we get

$$\tilde{T}(u)(t) = e^{tL}u^{in} + \int_0^t e^{(t-s)L} \left(f(u(s)) - Lu(s) \right) \mathrm{d}s.$$

- Looking at the derivative of $\tilde{T}$ at $\bar{u}$

$$\tilde{D}T(\bar{u})(h)(t) = \int_0^t e^{(t-s)L} \left(Df(\bar{u}(s)) - L \right) h(s) \mathrm{d}s,$$

we see that $\tilde{T}$ should be contracting if $L \approx Df(\bar{u}(s))$.

A different fixed point reformulation

$$\begin{cases} u'(t) - Lu(t) = f(u(t)) - Lu(t) & t \in [0, \tau] \\ u(0) = u^{in} \end{cases}$$

- Using Duhamel's principle/the variation of constants formula, we get

$$\tilde{T}(u)(t) = e^{tL} u^{in} + \int_0^t e^{(t-s)L} \left(f(u(s)) - Lu(s) \right) \mathrm{d}s.$$

- Looking at the derivative of $\tilde{T}$ at $\bar{u}$

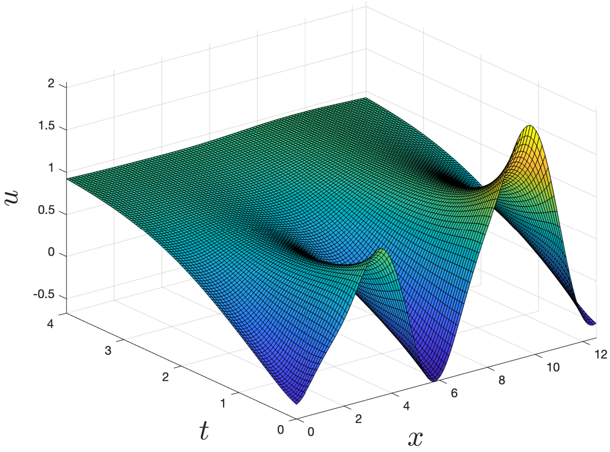
$$\tilde{D}T(\bar{u})(h)(t) = \int_0^t e^{(t-s)L} \left(Df(\bar{u}(s)) - L \right) h(s) \mathrm{d}s,$$

  we see that $\tilde{T}$ should be contracting if $L \approx Df(\bar{u}(s))$.

- We again split the time interval $0 = \tau_0 < \tau_1 < \ldots < \tau_M = \tau$, and take a different approximation on each smaller subinterval:

$$L^{(m)} \approx Df(\bar{u}^{(m)})(s), \qquad s \in [\tau_m, \tau_{m+1}].$$

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2} + u(1-u) & (t,x) \in (0,4] \times \mathbb{T}_{4\pi}, \\ u(0,\cdot) = u^{in}. \end{cases}$$



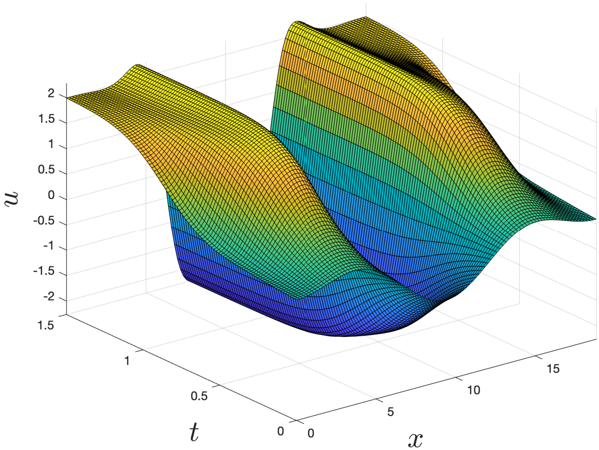
Theorem

$$\|\bar{u} - u\| \leq 5e^{-2}$$

$N = 14$
$K = 2$
$M = 25$

Application to parabolic PDEs 2: Swift-Hohenberg

$$\begin{cases} \frac{\partial u}{\partial t} = -\left(\frac{\partial^2}{\partial x^2} + 1\right)^2 u + 5u - u^3 & (t,x) \in (0, 1.5] \times \mathbb{T}_{6\pi}, \\ u(0,\cdot) = u^{in}. \end{cases}$$



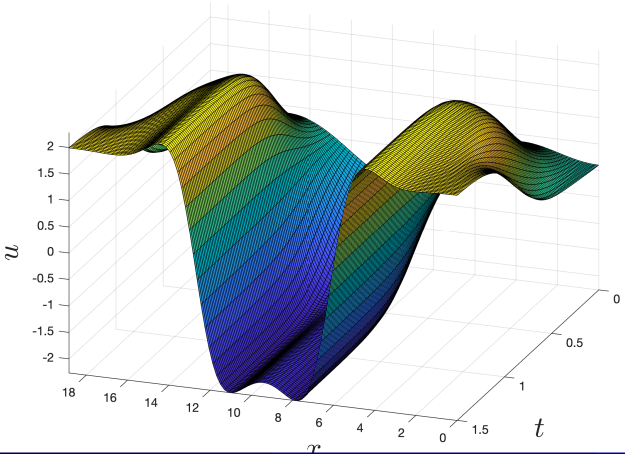
Theorem

$\|\bar{u} - u\| \leq 4e^{-8}$

$N = 30$
$K = 5$
$M = 100$

Application to parabolic PDEs 2: Swift-Hohenberg

$$\begin{cases} \dfrac{\partial u}{\partial t} = -\left(\dfrac{\partial^2}{\partial x^2}+1\right)^2 u + 5u - u^3 & (t,x)\in(0,1.5]\times\mathbb{T}_{6\pi}, \\ u(0,\cdot) = u^{in}. \end{cases}$$



Theorem

$\|\bar{u} - u\| \leq 4e^{-8}$

$N = 30$

$K = 5$

$M = 100$

THANK YOU FOR YOUR ATTENTION!