An introduction to computer-assisted proofs via a posteriori validation

Maxime Breden
CMAP, Ecole polytechnique

MAX team seminar, March 18, 2024
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.
Objective: prove *quantitative* theorems about some *specific* solutions of a given ODE or PDE, using numerical simulations.
Objective: prove \textit{quantitative} theorems about some \textit{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 \cite{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.
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.
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; ...].
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.
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

1. A simple example

2. Validated integration of ODEs using Chebyshev series

3. Alternate strategy
1. A simple example

2. Validated integration of ODEs using Chebyshev series

3. Alternate strategy
Motivation: periodic orbits and chaos

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

Maxime Breden

Computer-assisted proofs

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

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

We start by looking numerically for a period 3 orbit. To do so, we can consider the map \( F : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \) 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 \( F \)?

Maxime Breden
Computer-assisted proofs
MAX team seminar
We start by looking numerically for a period 3 orbit.

\[ x_{n+1} = \mu x_n (1 - x_n) \]
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.
- 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.
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 $F(\bar{X}) \approx 0$. 

\[ x_{n+1} = \mu x_n (1 - x_n) \]
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.
- 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!

\[ F : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \quad F(\bar{X}) \approx 0. \]

Theorem (à la Newton-Kantorovich)

Let \( \varepsilon, K, L > 0 \) such that

\[ \|F(\bar{X})\| \leq \varepsilon \|DF(\bar{X})^{-1}\| \leq \kappa \|DF(X) - DF(\bar{X})\| \leq L \|X - \bar{X}\| \quad \forall X \in \mathbb{R}^3. \]

If \( \varepsilon < \frac{1}{2} \kappa^2 L \),

then \( F \) has a unique zero \( X^* \) satisfying

\[ \|X^* - \bar{X}\| \leq r, \quad r = \frac{1 - \sqrt{1 - 2\kappa^2 L \varepsilon \kappa L}}{1}. \]

Proof:

\( T : \begin{array}{c} X \mapsto X - DF(\bar{X})^{-1}F(X) \end{array} \) is a contraction on the closed ball of center \( \bar{X} \) and radius \( r \).
We need proof!

\[ F : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \quad F(\bar{X}) \approx 0. \]

- We want to prove \textit{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

\[ \| F(\bar{X}) \| \leq \varepsilon \| \text{DF}(\bar{X})^{-1} \| \leq \kappa \| \text{DF}(X) - \text{DF}(\bar{X}) \| \leq L \| X - \bar{X} \| \quad \forall X \in \mathbb{R}^3. \]

If \( \varepsilon < \frac{1}{2} \kappa^2 L \), then \( F \) has a unique zero \( X^\ast \) satisfying

\[ \| X^\ast - \bar{X} \| \leq r, \quad r = \frac{1 - \sqrt{1 - 2 \kappa^2 \varepsilon \kappa L}}{\kappa L}. \]

Proof:

\[ T : X \mapsto X - \text{DF}(\bar{X})^{-1} F(X) \]

is a contraction on the closed ball of center \( \bar{X} \) and radius \( r \).
We need proof!

\[ F : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \quad F(\bar{X}) \approx 0. \]

- We want to prove \textbf{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

\[
\| F(\bar{X}) \| \leq \varepsilon \\
\| DF(\bar{X})^{-1} \| \leq \kappa \\
\| DF(X) - DF(\bar{X}) \| \leq L \| X - \bar{X} \| \quad \forall X \in \mathbb{R}^3.
\]

If

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

then \( F \) has a unique zero \( X^* \) satisfying \( \| X^* - \bar{X} \| \leq r, \quad r = \frac{1 - \sqrt{1 - 2\kappa^2 L \varepsilon}}{\kappa L}. \)
We need proof!

\[ F : \mathbb{R}^3 \to \mathbb{R}^3 \quad F(\bar{X}) \approx 0. \]

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

\[
\| F(\bar{X}) \| \leq \varepsilon
\]
\[
\| DF(\bar{X})^{-1} \| \leq \kappa
\]
\[
\| DF(X) - DF(\bar{X}) \| \leq L \| X - \bar{X} \| \quad \forall X \in \mathbb{R}^3.
\]

If

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

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

**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?

\[ g(a, b) = 333.75 b^6 + a^2 (11 a^2 b^2 - b^6 - 121 b^4 - 2) + 5.5 b^8 + a^2 b, \]
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 \( \epsilon \) that we numerically evaluated really bounds \( \| F(\bar{X}) \| \), or that \( \epsilon < \frac{1}{2} \kappa^2 L \)?
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 $\varepsilon$ that we numerically evaluated really bounds $\|F(\bar{X})\|$, or that $\varepsilon < \frac{1}{2\kappa^2 L}$?
Calling interval arithmetic to the rescue

Let $F$ be a set of floating point numbers, corresponding to the (finite!) set of real numbers that the computer can represent with a given precision, and $\bigtriangleup, \bigtriangledown : \mathbb{R} \to F$, the round-down and round-up operators.
Let $\mathbb{F}$ be a set of floating point numbers, corresponding to the (finite!) set of real numbers that the computer can represent with a given precision, and

$$\nabla, \Delta : \mathbb{R} \to \mathbb{F},$$

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

$$\bigtriangleup, \nabla : \mathbb{R} \rightarrow \mathbb{F},$$

the round-down and round-up operators.

Example: consider $x = 0.1$. In base 2, $x$ writes

$$x = (1.1001100110011001100...)_2 \times 2^{-4}.$$

With 8 bits of precision (for the mantissa), we have

$$\bigtriangleup(x) = (1.1001100)_2 \times 2^{-4} \quad \text{and} \quad \nabla (x) = (1.1001101)_2 \times 2^{-4}.$$
Let $\mathbb{F}$ be 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:

$$x \in \mathbb{R} \quad \to \quad [x] := [\nabla(x), \triangle(x)].$$
Calling interval arithmetic to the rescue

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

$$\nabla, \triangle : \mathbb{R} \rightarrow \mathbb{F},$$

the round-down and round-up operators.

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

$$x \in \mathbb{R} \rightarrow [x] := [\nabla(x), \triangle(x)].$$

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

$$x + y \rightarrow [x] [+] [y],$$

where $[+]$ is defined as follows

$$[x] [+] [y] := [\nabla(\nabla(x) + \nabla(y)) , \triangle(\triangle(x) + \triangle(y))].$$
Let $\mathbb{F}$ be a set of floating point numbers, corresponding to the (finite!) set of real numbers that the computer can represent with a given precision, and

$$\nabla, \triangle : \mathbb{R} \rightarrow \mathbb{F},$$

the round-down and round-up operators.

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

$$x \in \mathbb{R} \rightarrow [x] := [\nabla(x), \triangle(x)].$$

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

$$x + y \rightarrow [x] [+][y],$$

where $[+]$ is defined as follows (doable in practice, IEEE 754 standard)

$$[x] [+][y] := [\nabla(\nabla(x) + \nabla(y)), \triangle(\triangle(x) + \triangle(y))].$$
Let $\mathbb{F}$ be a set of floating point numbers, corresponding to the (finite!) set of real numbers that the computer can represent with a given precision, and

$$\nabla, \triangle : \mathbb{R} \rightarrow \mathbb{F},$$

the round-down and round-up operators.

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

$$x \in \mathbb{R} \rightarrow [x] := [\nabla(x), \triangle(x)].$$

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

$$x + y \rightarrow [x] [+][y],$$

where $[+]$ is defined as follows (doable in practice, IEEE 754 standard)

$$[x] [+][y] := [\nabla(\nabla(x) + \nabla(y)), \triangle(\triangle(x) + \triangle(y))].$$

We then have $x + y \in [x] [+][y]$. 

We reformulate the search of a period 3 orbit as a zero-finding problem

\[ F(X) = 0. \]

We numerically find an approximate solution.

We estimate a posteriori \( \|F(\bar{X})\|, \|DF(\bar{X})^{-1}\| \) and \( \|D^2F(X)\| \), and do so rigorously using interval arithmetic.

We use these estimates to prove that \( T: X \mapsto X - DF(\bar{X})^{-1} F(X) \) is a contraction on a small neighborhood of \( \bar{X} \).
We reformulate the search of a period 3 orbit as a zero-finding problem $F(X) = 0$.
We reformulate the search of a period 3 orbit as a zero-finding problem \( F(X) = 0 \).

We numerically find an approximate solution.
\[ x_{n+1} = \mu x_n (1 - x_n) \]

1. We reformulate the search of a period 3 orbit as a zero-finding problem \( F(X) = 0 \).
2. We numerically find an approximate solution.
3. We estimate a posteriori

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

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 \( F(X) = 0 \).

2. We numerically find an approximate solution.

3. We estimate a posteriori

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

and do so rigorously using interval arithmetic.

4. We use these estimates to prove that

\[ T : X \mapsto X - DF(\bar{X})^{-1} F(X) \]

is a contraction on a small neighborhood of \( \bar{X} \).
Outline

1. A simple example
2. Validated integration of ODEs using Chebyshev series
3. Alternate strategy
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 choose a Banach space $X$, and in particular a norm.

2. Find numerically an approximate zero $\bar{X}$.

   ▶ Choice of discretization method, of a finite dimensional space $X_h$ in which we look for the approximate solution.

3. Estimate a posteriori $\|F(\bar{X})\|\|DF(\bar{X})^{-1}\|$ and $\|D^2F(X)\|$.

   ▶ The main difficulty lies in controlling $\|DF(\bar{X})^{-1}\|$.
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 choose a Banach space $X$, 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 choose a Banach space \( X \), and in particular a norm.

2. Find numerically an approximate zero \( \bar{X} \).
   - Choice of discretization method, of a finite dimensional space \( X_h \) 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 choose a Banach space $X$, and in particular a norm.

2. Find numerically an approximate zero $\bar{X}$.
   - Choice of discretization method, of a finite dimensional space $X_h$ 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 choose 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
   
   $\|F(\bar{X})\|$, $\|DF(\bar{X})^{-1}\|$ and $\|D^2F(X)\|$.

   - 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

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

Maxime Breden
Computer-assisted proofs
MAX team seminar
Setting for validated integration of ODEs

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

with \( f : \mathbb{R}^d \to \mathbb{R}^d \) smooth and \( \tau > 0 \) fixed.
Setting for validated integration of ODEs

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

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

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

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:**

\[ F(u)(t) = u(t) - \left( u^{in} + \int_0^t f(u(s)) \, ds \right) . \]
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 \( f : \mathbb{R}^d \rightarrow \mathbb{R}^d \) smooth and \( \tau > 0 \) fixed.

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

- **Main idea for the zero-finding problem:**

  \[
  F(u)(t) = u(t) - \left( u^{in} + \int_0^t f(u(s))ds \right).
  \]

- **Key observation:**

  \[
  DF(\bar{u})(h)(t) = h(t) - \int_0^t Df(\bar{u}(s))h(s)ds,
  \]

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

\[
\begin{aligned}
\left\{
\begin{array}{ll}
u'(t) = \tau f(u(t)) & t \in [-1, 1] \\
u(-1) = u^{in}
\end{array}
\right.
\end{aligned}
\]
Look for the solution as a Chebyshev series:

\[ u(t) = u_0 + 2 \sum_{n=1}^{\infty} u_n T_n(t), \quad T_n(\cos \theta) = \cos(n\theta). \]
\[
\begin{aligned}
&\begin{cases}
  u'(t) = \tau f(u(t)) & t \in [-1, 1] \\
  u(-1) = u^{in}
\end{cases} \\
&\text{Look for the solution as a Chebyshev series:}
\end{aligned}
\]

\[
\begin{aligned}
u(t) &= u_0 + 2 \sum_{n=1}^{\infty} u_n T_n(t), \\
&\quad T_n(\cos \theta) = \cos(n\theta).
\end{aligned}
\]

\[
\begin{aligned}
&\text{The unknown is the sequence } u = (u_n)_{n \geq 0} \text{ of Chebyshev coefficients.}
\end{aligned}
\]
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:

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

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

By plugging the Chebyshev series ansatz into

\[ u(t) - \left( u^{in} + \tau \int_{-1}^{t} f(u(s)) \, ds \right) = 0, \]

we obtain our \( F(u) = 0 \) problem.
Chebyshev series

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

- Look for the solution as a Chebyshev series:

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

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

- By plugging the Chebyshev series ansatz into

\[
u(t) - \left( u^\text{in} + \tau \int_{-1}^{t} f(u(s)) \, ds \right) = 0,
\]

we obtain our \( F(u) = 0 \) problem.

- The approximate solution \( \bar{u} \) 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:

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

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

▸ By plugging the Chebyshev series ansatz into

\[
u(t) - \left( u^{in} + \tau \int_{-1}^{t} f(u(s))ds \right) = 0,
\]

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

▸ The approximate solution \( \bar{u} \) is taken as a truncated Chebyshev series.

▸ We look for the exact solution in the space \( \ell^1_\nu := \{ \mathbf{u}, \| \mathbf{u} \|_\nu < \infty \}, \)

\[
\| \mathbf{u} \|_\nu := |u_0| + 2 \sum_{n=1}^{\infty} |u_n| \nu^n, \quad \nu \geq 1.
\]
Why Chebyshev series?

\[ u(t) - \left( u^\text{in} + \tau \int_{-1}^{t} f(u(s)) \, ds \right) = 0. \]
Why Chebyshev series?

\[ u(t) - \left( u^{in} + \tau \int_{-1}^{t} f(u(s)) \, ds \right) = 0. \]

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

\[ T_n = \frac{1}{2} \left( \frac{1}{n} + \frac{1}{T_n} + \frac{1}{1} - \frac{1}{n} - \frac{1}{T_n} \right). \]

- Efficient computations of nonlinearities using the FFT.

- Computing \( \|F(\overline{u})\|_\nu \) is rather straightforward.

- \( \ell_1 \nu \) is a Banach algebra:
  \[ \|u \circledast v\|_\nu \leq \|u\|_\nu \|v\|_\nu. \]

Simplifies the estimation of \( \|D^2 F(u)\|_\nu \) for \( u \) in a neighborhood of \( \overline{u} \).
Why Chebyshev series?

\[ u(t) = \left( u^\text{in} + \tau \int_{-1}^{t} f(u(s)) \, ds \right) = 0. \]

- 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?

\[ u(t) - \left( u^{in} + \tau \int_{-1}^{t} f(u(s)) ds \right) = 0. \]

- 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?

\[ u(t) - \left( u^{in} + \tau \int_{-1}^{t} f(u(s))ds \right) = 0. \]

- 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 \( \| F(\bar{u}) \|_\nu \) is rather straightforward.
Why Chebyshev series?

\[ u(t) - \left( u^{in} + \tau \int_{-1}^{t} f(u(s))ds \right) = 0. \]

- 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 \( \|F(\bar{u})\|_\nu \) is rather straightforward.

- \( \ell^1_\nu \) is a Banach algebra: \( \|u * v\|_\nu \leq \|u\|_\nu \|v\|_\nu \).
Why Chebyshev series?

\[ u(t) - \left( u^{in} + \tau \int_{-1}^{t} f(u(s)) \, ds \right) = 0. \]

- 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 \( \|F(\bar{u})\|_\nu \) is rather straightforward.

- \( \ell^1_\nu \) is a Banach algebra: \( \|u * v\|_\nu \leq \|u\|_\nu \|v\|_\nu \).
  - Simplifies the estimation of \( \|D^2F(u)\|_\nu \) for \( u \) in a neighborhood of \( \bar{u} \).
How to construct the approximate inverse $A$
How to construct the approximate inverse $\tilde{A}$

$$DF(\bar{u}) =$$
How to construct the approximate inverse $A$

$DF(\bar{u}) \approx $
How to construct the approximate inverse $A$
Using this constructing, when keeping the first $N$ Chebyshev modes in the finite block, we get
\[ \| I - ADF(\bar{u}) \|_\nu \approx \tau \| f'(\bar{u}) \|_\nu N. \]

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

[Lessard Reinhardt '14]
Maxime Breden
Computer-assisted proofs
MAX team seminar
Using this constructing, when keeping the first $N$ Chebyshev modes in the finite block, we get

$$\| I - A D F(\bar{u}) \|_\nu \approx \tau \frac{\| f'(\bar{u}) \|_\nu}{N}.$$
Using this constructing, when keeping the first $N$ Chebyshev modes in the finite block, we get

$$\|I - A F(\bar{u})\|_\nu \approx \frac{\tau \|f'(\bar{u})\|_\nu}{N}.$$ 

Up to taking $N$ large enough, we can therefore get $\|I - A F(\bar{u})\|_\nu < 1$, and hope to apply the entire \textit{a posteriori validation} procedure.
Using this constructing, when keeping the first $N$ Chebyshev modes in the finite block, we get

$$\| I - A D F(\bar{u}) \|_\nu \approx \frac{\tau \| f'(\bar{u}) \|_\nu}{N}.$$ 

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

[Lessard Reinhardt ’14]
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 $[\tau_{m-1}, \tau_m]$: $u(1)(t) - (u(1)(0) + \int_0^t f(u(1)(s)) \, ds) = 0$ for $t \in [0, \tau_1]$, $u(2)(t) - (u(1)(\tau_1) + \int_{\tau_1}^t f(u(2)(s)) \, ds) = 0$ for $t \in [\tau_1, \tau_2]$, ...

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

[van den Berg Sheombarsing ‘21] Maxime Breden

Computer-assisted proofs

MAX team seminar
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 $[\tau_{m-1}, \tau_m]$:

\[
u^{(1)}(t) - \left( u^{in} + \int_0^t f(u^{(1)}(s))ds \right) = 0 \quad t \in [0, \tau_1],
\]

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

\[\vdots\]

\[
u^{(M)}(t) - \left( u^{(M-1)}(\tau_{M-1}) + \int_{\tau_{M-1}}^t f(u^{(M)}(s))ds \right) = 0 \quad t \in [\tau_{M-1}, \tau].
\]
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 $[\tau_{m-1}, \tau_m]$:

\[
\begin{align*}
    u^{(1)}(t) - \left( u^{in} + \int_0^t f(u^{(1)}(s))ds \right) &= 0 & \quad t \in [0, \tau_1], \\
    u^{(2)}(t) - \left( u^{(1)}(\tau_1) + \int_{\tau_1}^t f(u^{(2)}(s))ds \right) &= 0 & \quad t \in [\tau_1, \tau_2], \\
    \vdots \\
    u^{(M)}(t) - \left( u^{(M-1)}(\tau_{M-1}) + \int_{\tau_{M-1}}^t f(u^{(M)}(s))ds \right) &= 0 & \quad t \in [\tau_{M-1}, \tau].
\end{align*}
\]

- Each $u^{(m)}$ is then represented by a Chebyshev series, and this leads to a big $F(u) = 0$ 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 \([\tau_{m-1}, \tau_m]\):

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

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

- [van den Berg Sheombarising '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]

\[ x' = 10(x - y) \]
\[ y' = 28x - y - xz \]
\[ z' = -\frac{8z}{3} + xy \]

Integration time \( \tau \approx 100 \)
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].
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 Mezarobba '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].
1. A simple example
2. Validated integration of ODEs using Chebyshev series
3. Alternate strategy
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

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

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

\[
F(u)(t) = u(t) - \left( u^{\text{in}} + \int_0^t f(u(s))\,ds \right),
\]

and then into a fixed point problem \( T(u) = u - AF(u) \).
A different fixed point reformulation

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

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

\[
F(u)(t) = u(t) - \left( u^{in} + \int_0^t f(u(s))\,ds \right),
\]

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

- One could also directly get a fixed point problem:

\[
\tilde{T}(u)(t) = u^{in} + \int_0^t f(u(s))\,ds.
\]
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:

\[
F(u)(t) = u(t) - \left( u^{in} + \int_0^t f(u(s)) \, ds \right),
\]

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

One could also directly get a fixed point problem:

\[
\tilde{T}(u)(t) = u^{in} + \int_0^t f(u(s)) \, ds.
\]

\( \tilde{T} \) has no reason to be contracting near \( \bar{u} \), except for \( \tau \) small.
A different fixed point reformulation

\[
\begin{cases}
    u'(t) - Lu(t) = f(u(t)) - Lu(t) \quad t \in [0, \tau] \\
    u(0) = u^{in}
\end{cases}
\]
A different fixed point reformulation

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

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

\[\tilde{T}(u)(t) = e^{tL}u^{in} + \int_0^t e^{(t-s)L} (f(u(s)) - Lu(s)) \, ds.\]
A different fixed point reformulation

\[
\begin{aligned}
&\left\{
\begin{array}{ll}
u'(t) - Lu(t) = f(u(t)) - Lu(t) & t \in [0, \tau] \\
u(0) = u^{in}
\end{array}
\right.
\end{aligned}
\]

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

\[
\tilde{T}(u)(t) = e^{tL}u^{in} + \int_0^t e^{(t-s)L}(f(u(s)) - Lu(s)) \, ds.
\]

Looking at the derivative of \( \tilde{T} \) at \( \bar{u} \)

\[
\tilde{D}T(\bar{u})(h)(t) = \int_0^t e^{(t-s)L}(Df(\bar{u}(s)) - L) h(s) \, ds,
\]

we see that \( \tilde{T} \) should be contracting if \( L \approx Df(\bar{u}(s)) \).
A different fixed point reformulation

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

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

\[
\tilde{T}(u)(t) = e^{tL}u^{in} + \int_{0}^{t} e^{(t-s)L}(f(u(s)) - Lu(s))\,ds.
\]

- Looking at the derivative of \(\tilde{T}\) at \(\tilde{u}\)

\[
\tilde{D}T(\tilde{u})(h)(t) = \int_{0}^{t} e^{(t-s)L}(Df(\tilde{u}(s)) - L)h(s)\,ds,
\]

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

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

\[
L^{(m)} \approx Df(\tilde{u}^{(m)})(s), \quad s \in [\tau_m, \tau_{m+1}].
\]
Application to parabolic PDEs 1: Fisher-KPP

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

Theorem

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

- $N = 14$
- $K = 2$
- $M = 25$
Theorem
\[ \| \tilde{u} - u \| \leq 4e^{-8} \]

\[ N = 30 \]
\[ K = 5 \]
\[ M = 100 \]
\[ \begin{aligned}
\frac{\partial u}{\partial t} &= -\left(\frac{\partial^2}{\partial x^2} + 1\right)^2 u + 5u - u^3 & (t, x) \in (0, 1.5] \times \mathbb{T}_{6\pi}, \\
u(0, \cdot) &= u^{in}.
\end{aligned} \]
THANK YOU FOR YOUR ATTENTION!