

Operator-Adapted Initial-Data Extensions for Hard-Constrained Neural PDE Solvers

Rémy Hosseinkhan-Boucher, Nicolas Baradel, Charles-Albert Lehalle

14 October 2026

Groupe de travail MathsFi — Centre de Mathématiques Appliquées (CMAP), École polytechnique



Intention and Chain of Reasoning

Intention (why the question matters)

- gap $\rightarrow$ under-developed methodology, inferred from literature, conferences, discussions
- precision and rigour $\rightarrow$ projection of the state of the art onto the discipline (mathematics)
- applications in finance

Chain of reasoning (main steps to solve the problem)

1. problem, question
2. naive solution, on an example
3. pathology (differentiability)
4. isolate the pathology, minimal example
5. bibliography (knowledge base)
6. proof of concept
7. formalisation and understanding, against the knowledge base

exact initial condition $\rightarrow$ which fixed field h ?

$h^0 = g$ (Lagaris et al., 1998)

$g \notin C^{2p} \rightarrow$ Dirac masses, $J(\theta) = +\infty$

butterfly payoff, heat evolution

C_0 -semigroups (Arendt et al., 2011), Sobolev spaces, spectral bias (Rahaman et al., 2019)

trained errors, three test cases

own knowledge, books of a personal library $\rightarrow$
Propositions 4.1 to 4.3

Artificial Neural Networks for Solving Ordinary and Partial Differential Equations

Isaac Elias Lagaris, Aristidis Likas, *Member, IEEE*, and Dimitrios I. Fotiadis

We choose a form for the trial function $\Psi_t(\vec{x})$ such that by construction satisfies the BC's. This is achieved by writing it as a sum of two terms

$$\Psi_t(\vec{x}) = A(\vec{x}) + F(\vec{x}, N(\vec{x}, \vec{p})) \quad (4)$$

where $N(\vec{x}, \vec{p})$ is a single-output feedforward neural network with parameters $\vec{p}$ and n input units fed with the input vector $\vec{x}$.

The term $A(\vec{x})$ contains no adjustable parameters and satisfies the boundary conditions.

Two terms

$A \rightarrow$ boundary conditions, no weights

$F \rightarrow$ network N , no boundary contribution

Consequence

boundary conditions exact, independent of $\vec{p}$

training $\rightarrow$ differential equation only

A inside the domain $\rightarrow$ unspecified

Section II, page 988

same construction $\rightarrow$ distance functions, transfinite interpolation (Sukumar and Srivastava, 2022)

The Initial-Value Problem

Setting

- $\Omega \in \{\mathbb{T}, \mathbb{R}\}$, $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$, $Q = \Omega \times (0, T)$
- $\mathcal{L}^X = \sum_{j=0}^{2p} \gamma_{\mathcal{L}^X, j} \frac{\partial^j}{\partial x^j}$ uniformly parabolic
- $g \in C^0(\Omega; \mathbb{R}) \cap L^2(\Omega; \mathbb{R}) \rightarrow$ first-derivative discontinuity allowed

$$\mathcal{L}u^* = \partial_t u^* - \mathcal{L}^X u^* = 0 \text{ on } Q, \quad u^*(x, 0) = g(x)$$

Example 3.1 – butterfly payoff

$$\Omega = \mathbb{R}, \quad p = 1; \quad \nu, \ell \in (0, +\infty); \quad \mu, \varrho, x^* \in \mathbb{R}$$

$$\mathcal{L}^X = \nu \partial_{xx}^2 + \mu \partial_x - \varrho$$

operator

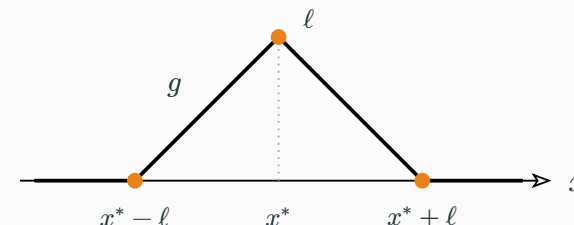
$$g(x) = (\ell - |x - x^*|)^+$$

payoff

$$\partial_{xx}^2 g = \delta_{x^* - \ell} - 2\delta_{x^*} + \delta_{x^* + \ell}$$

second distributional derivative

- $g \notin C^1(\mathbb{R}; \mathbb{R}) \rightarrow$ first-derivative jumps 1, -2, 1 (orange)
- compactly supported
- Lipschitz



The Constrained Model Class

Objective

$$\Phi_{\theta}^h(x, t) = h(x, t) + d_0(t) \Psi_{\theta}(x, t) \quad \text{learns } u^*(x, t)$$

- $h \longrightarrow$ candidate field, fixed before training
- $\Psi_{\theta} \longrightarrow$ neural field
- $d_0(0) = 0, d_0(t) > 0$ for $t \in (0, T]$

Two requirements on h

trace $h(x, 0) = g(x) \longrightarrow \Phi_{\theta}^h(x, 0) = g(x) \quad \forall \theta$ (initial condition exact)

regularity $h \in \mathcal{W} \longrightarrow \mathcal{L}h$ well defined pointwise, $\mathcal{L}h \in C^0(Q)$

$$\mathcal{W} = \left\{ v \in C^0(\overline{Q}) : \partial_t v, \frac{\partial^j v}{\partial x^j} \in C^0(Q), 1 \leq j \leq 2p \right\}$$

Residual of the candidate field

$$\mathcal{L}\Phi_{\theta}^h = \underbrace{\mathcal{L}h}_{\text{fixed}} + \underbrace{\mathcal{N}_{\theta}}_{\text{trained}}, \quad \mathcal{N}_{\theta} = d_0 \mathcal{L}\Psi_{\theta} + d_0' \Psi_{\theta}$$

Fitting Objective and Forcing Energy

Fitting objective

$$\min_{\theta \in \Theta} J(\theta), \quad J(\theta) = \|\mathcal{L}\Phi_{\theta}^h\|_{L^2(Q)}^2 = \int_0^T \int_{\Omega} |\mathcal{L}\Phi_{\theta}^h(x, t)|^2 \, dx \, dt$$

Correction problem

$$\mathcal{L}\Phi_{\theta}^h = 0 \iff \mathcal{L}(d_0\Psi_{\theta}) = -\mathcal{L}h \text{ on } Q, \quad (d_0\Psi_{\theta})(x, 0) = 0$$

Forcing energy

$E_h = \|\mathcal{L}h\|_{L^2(Q)}^2 \rightarrow$ energy of the source $-\mathcal{L}h$, fixed before training

Expansion of the objective

$$E_h < +\infty, \mathcal{N}_{\theta} \in L^2(Q) \implies J(\theta) = E_h + 2\langle \mathcal{L}h, \mathcal{N}_{\theta} \rangle_{L^2(Q)} + \|\mathcal{N}_{\theta}\|_{L^2(Q)}^2$$

- $\mathcal{N}_{\theta}$ approximates $-\mathcal{L}h$
- cross term $2\langle \mathcal{L}h, \mathcal{N}_{\theta} \rangle_{L^2(Q)}$ of either sign
- $J(\theta)$ may lie below E_h

The Time-Independent Field h^0 : Infinite Objective

Field held constant in time

$h^0(x, t) = g(x) \rightarrow$ construction of Lagaris et al. (1998)

$g \notin C^{2p} \rightarrow h^0 \notin \mathcal{W} \rightarrow \mathcal{L}h^0$ undefined classically

Residual of h^0 for the butterfly payoff

bounded part + three **Dirac masses** $\rightarrow$ not sampled

$$\mathcal{L}h^0 = \underbrace{-\mu\partial_x g + \varrho g}_{\text{bounded}} - \nu(\delta_{x^*-\ell} - 2\delta_{x^*} + \delta_{x^*+\ell})$$

Objective J , infinite

$$\mathcal{N}_\theta \in L^2(Q) \implies J(\theta) = \|\mathcal{L}\Phi_\theta^{h^0}\|_{L^2(Q)}^2 = +\infty$$

Sampled residual $\mathcal{R}_\mathcal{D}$, finite

$$\begin{aligned} \mathcal{R}_\mathcal{D}(\theta) &= \frac{1}{|\mathcal{D}|} \sum_{z \in \mathcal{D}} |\mathcal{L}\Phi_\theta^{h^0}(z)|^2 \\ &= \frac{1}{|\mathcal{D}|} \sum_{z \in \mathcal{D}} |\mathcal{L}h^0(z)|^2 < +\infty \quad \text{at } \Psi_\theta = 0 \end{aligned}$$

Smoothed field $h_\varepsilon^0(x, t) = (g * G_\varepsilon)(x)$

expected sampled residual **diverges** as $\varepsilon \rightarrow 0$

$\mathcal{R}_\mathcal{D}(\theta) < +\infty \quad \text{vs.} \quad J(\theta) = +\infty$

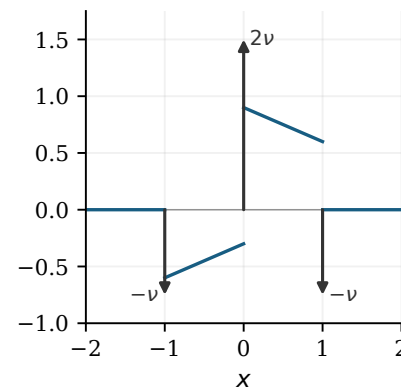


Figure 1(a): $\mathcal{L}h^0$ in x
curve $\rightarrow$ bounded part
arrows $\rightarrow$ Dirac masses

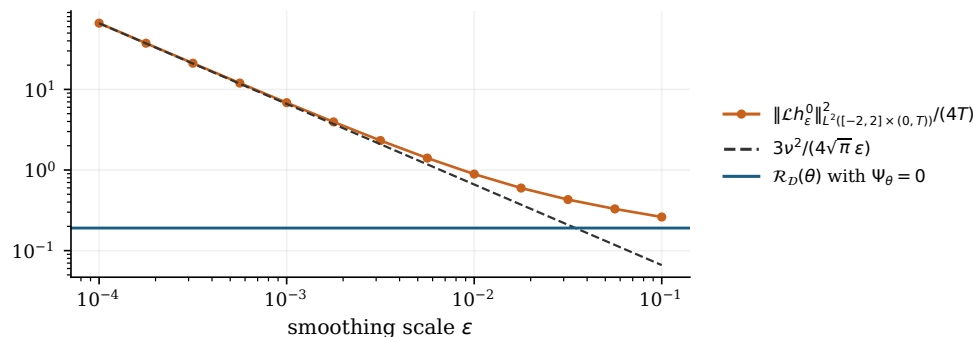


Figure 1(d): smoothed field h_ε^0 , width ε
orange $\rightarrow$ mean sampled residual, grows like $1/\varepsilon$ (dashed)
horizontal $\rightarrow$ sampled residual of h^0 , finite

From a Linear ODE to the Semigroup

Splitting of the operator

$$\mathcal{L}^X = A + B \quad A \rightarrow \text{retained, } e^{tA} \text{ computable} \quad B \rightarrow \text{remainder}$$

Linear ODE in $\mathbb{R}^d$

$$u' = Mu, \quad u(0) = u_0 \implies u(t) = e^{tM}u_0, \quad M \in \mathbb{R}^{d \times d}$$

Operator exponential

$$\partial_t u = \mathcal{L}^X u, \quad u(x, 0) = g(x) \implies u^*(\cdot, t) = e^{t\mathcal{L}^X} g$$

constant coefficients $\rightarrow$ Fourier diagonalises $\mathcal{L}^X \rightarrow \hat{u}^*(\xi, t) = e^{t\lambda_{\mathcal{L}^X, \xi}} \hat{g}(\xi)$, as e^{tM} for a diagonal M

Here: semigroup of A

$$h(x, t) = (e^{tA}g)(x) \implies \mathcal{L}h = -Bh, \quad u^*(\cdot, t) = e^{tB}h(\cdot, t) \quad (\text{constant coefficients})$$

- Goal** $\rightarrow$ high frequencies of u^* encoded in h , so that Ψ_θ does not have to learn them
- $\rightarrow$ networks fit low frequencies first (Rahaman et al., 2019)
 - $\rightarrow$ Fourier coefficients of $u^* - h$ bounded by $C |\xi|^{m_B - 2p} |\hat{g}(\xi)|$ for $|\xi| \geq R$ (Proposition 4.5)

Choices of the Fixed Field h

Retained evolution

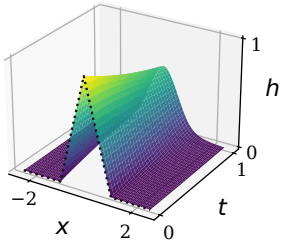
$\partial_t h = Ah$ on Q , $h(x, 0) = g(x)$, $B = \mathcal{L}^X - A$ of order m_B test cases $\longrightarrow \mathcal{L}^X = \gamma_{2p} \partial^{2p} / \partial x^{2p} + \gamma_1 \partial / \partial x + \gamma_0$

Retained operator A	Remainder B	m_B	Field $h(x, t) = (e^{tA} g)(x)$
0	$\gamma_{2p} \partial^{2p} / \partial x^{2p} + \gamma_1 \partial / \partial x + \gamma_0$	$2p$	$g(x)$ (Lagaris et al., 1998)
$\gamma_1 \partial / \partial x + \gamma_0$	$\gamma_{2p} \partial^{2p} / \partial x^{2p}$	$2p$	$e^{\gamma_0 t} g(x + \gamma_1 t)$
$\gamma_{2p} \partial^{2p} / \partial x^{2p}$	$\gamma_1 \partial / \partial x + \gamma_0$	1	$(G_t * g)(x)$, $\hat{G}_t(\xi) = e^{t\gamma_{2p}(i\xi)^{2p}}$
$\gamma_{2p} \partial^{2p} / \partial x^{2p} + \gamma_1 \partial / \partial x$	γ_0	0	$e^{-\gamma_0 t} u^*(x, t)$
$\mathcal{L}^X$	0	0	$u^*(x, t)$ (control)

Chen–Mangasarian smoothing

$$h(x, t) = \left(K_{\varepsilon(t)} * g\right)(x), \quad K_{\varepsilon}(y) = \frac{\varepsilon^2}{2(y^2 + \varepsilon^2)^{3/2}}, \quad \varepsilon(t) \propto t$$

call payoff $(s - K)^+ \longrightarrow \frac{1}{2} \left(s - K + \sqrt{(s - K)^2 + \varepsilon^2}\right)$
 $\longrightarrow$ fixed term of Zhang, Guo and Lu (2026), other width

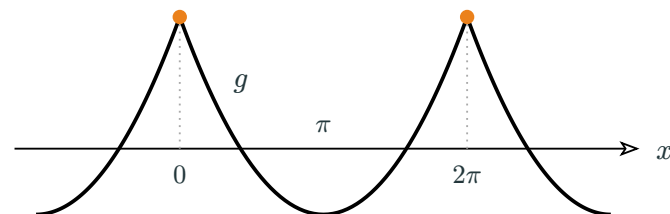


Question

$\longrightarrow$ which h gives the smallest solution error, at fixed network and training budget?

Periodic datum, $\Omega = \mathbb{T}$

- first-derivative discontinuity at 0
- $g \in H^1(\mathbb{T}) \setminus H^2(\mathbb{T})$
- training datum $\longrightarrow g$ itself, no Fourier projection



Two operators on $\mathbb{T}$

$$\partial_t u = \mathcal{L}^X u \text{ on } \mathbb{T} \times (0, 1), \quad u(x, 0) = g(x)$$

$$\text{Black-Scholes} \quad \longrightarrow \quad \mathcal{L}^X = \frac{1}{2}\sigma^2 \partial_{xx}^2 + \left(r - \frac{1}{2}\sigma^2\right) \partial_x - r \quad (\text{log-price, } \sigma = 0.5, r = 0.03)$$

$$\text{Order four} \quad \longrightarrow \quad \mathcal{L}^X = \gamma_{\mathcal{L}^X, 4} \partial_{xxxx}^4 + \gamma_{\mathcal{L}^X, 1} \partial_x + \gamma_{\mathcal{L}^X, 0}, \quad \gamma_{\mathcal{L}^X, 4} < 0$$

Butterfly on $\mathbb{R}$

$$\partial_t u = \nu \partial_{xx}^2 u + \mu \partial_x u - \varrho u \text{ on } \mathbb{R} \times (0, 1), \quad u(x, 0) = (\ell - |x - x^*|)^+$$

Results: Solution Error by Choice of h

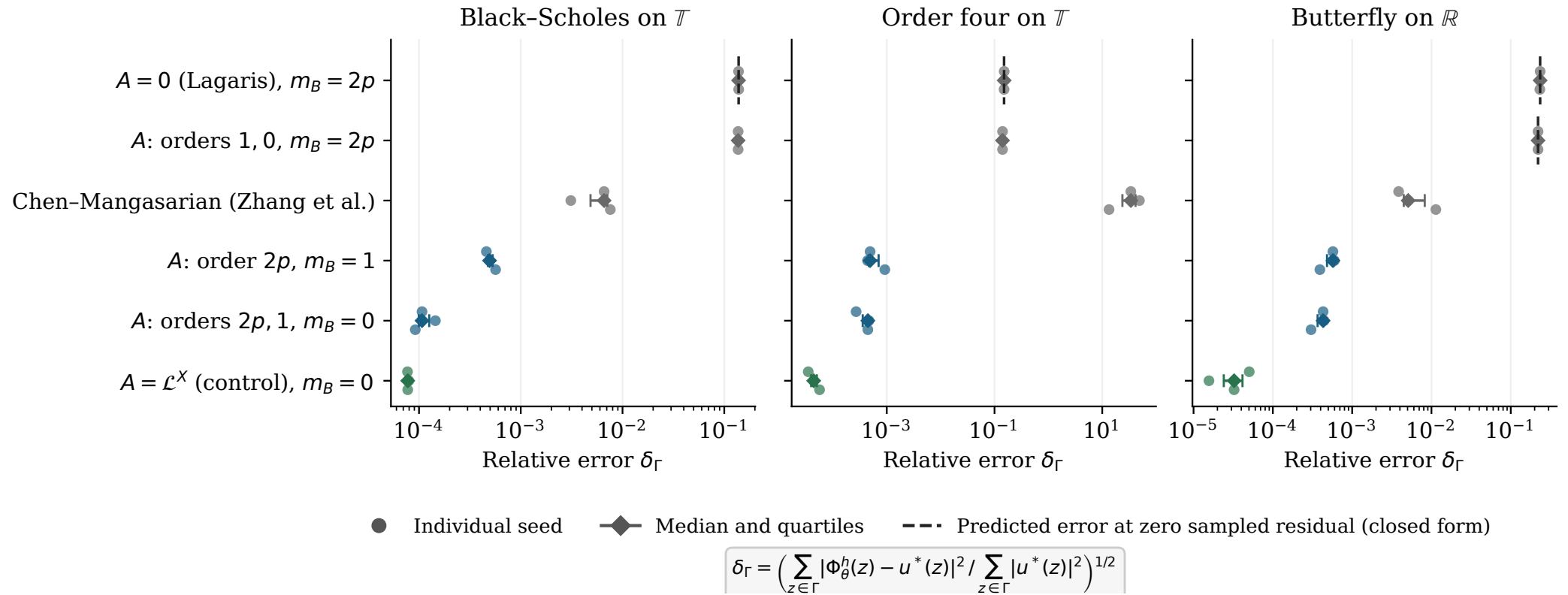


Figure 4 of the paper, rows of this talk, full datum g on $\mathbb{T}$
blue $\rightarrow m_B \leq p$, **green** $\rightarrow$ control, **grey** $\rightarrow$ other choices of h

Median Solution Error by Choice of h

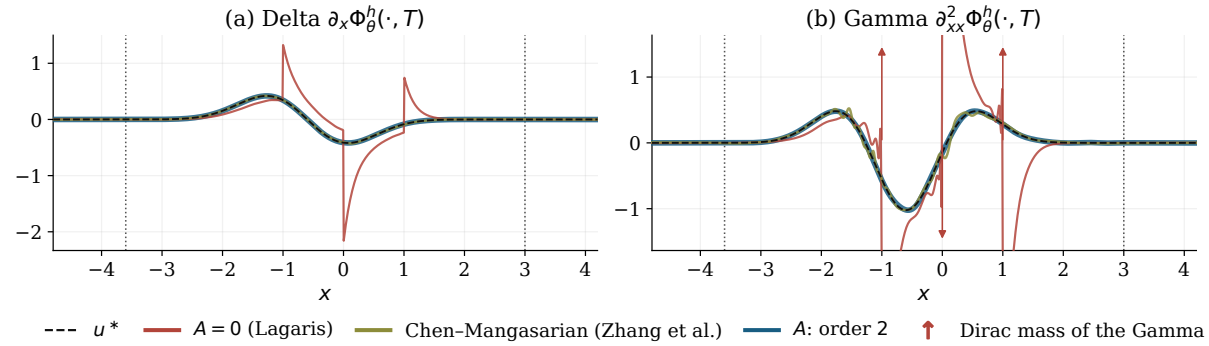
Median relative error $\tilde{\delta}_h$ (measured, 3 seeds, rounded to three digits)

Choice of h	Black-Scholes on $\mathbb{T}$	Order four on $\mathbb{T}$	Butterfly on $\mathbb{R}$
$A = 0$ (Lagaris), $m_B = 2p$	0.138	0.151	0.235
A : orders 1, 0, $m_B = 2p$	0.137	0.141	0.221
Chen-Mangasarian (Zhang et al.)	6.59×10^{-3}	33.9	5.09×10^{-3}
A : order $2p$, $m_B = 1$	4.93×10^{-4}	4.90×10^{-4}	5.72×10^{-4}
A : orders $2p, 1$, $m_B = 0$	1.07×10^{-4}	4.45×10^{-4}	4.30×10^{-4}
$A = \mathcal{L}^X$ (control), $m_B = 0$	7.75×10^{-5}	4.40×10^{-5}	3.25×10^{-5}

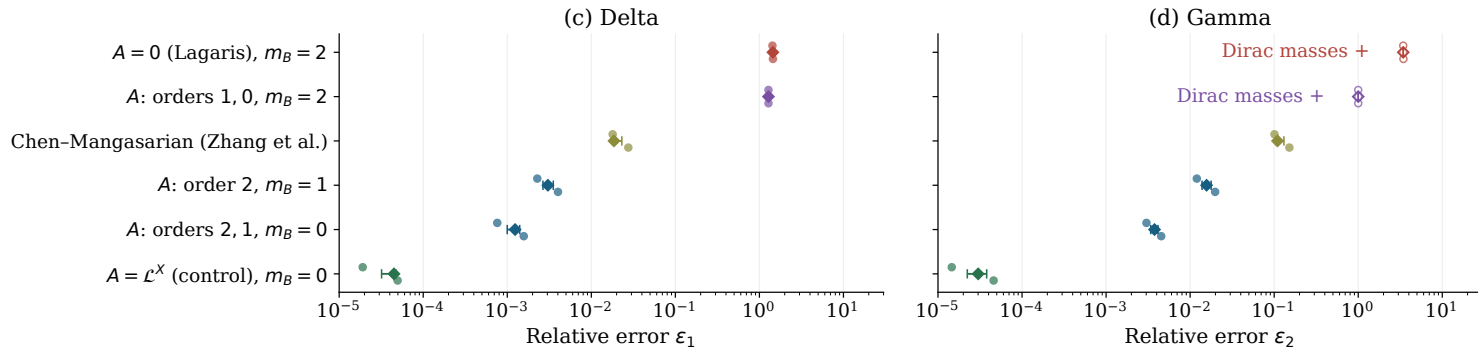
$m_B \leq p \rightarrow$ **smallest median error** on the three test cases, control excluded

$m_B = 2p \rightarrow$ median error larger than A : order $2p$ by a factor $\simeq 2.8 \times 10^2$ to $\simeq 4.1 \times 10^2$

Results: Greeks of the Butterfly on $\mathbb{R}$ at $t = T$



seed attaining the median error δ_T ; arrows $\rightarrow$ Dirac masses of $\partial_{xx}^2 h(\cdot, T)$ for $A=0$; dotted $\rightarrow$ interior window



3 seeds, median and quartiles; $\varepsilon_j = \left(\sum_{x \in I} |\partial_x^j \Phi_\theta^h(x, T) - \partial_x^j u^*(x, T)|^2 / \sum_{x \in I} |\partial_x^j u^*(x, T)|^2 \right)^{1/2}$, $j \in \{1, 2\}$, I = grid points of the interior window; hollow $\rightarrow$ classical part only, the Gamma having Dirac masses

$m_B \leq p \rightarrow \varepsilon_1 \leq 3.1 \times 10^{-3}, \varepsilon_2 \leq 1.6 \times 10^{-2}$ (measured, median of 3 seeds)
 Chen-Mangasarian $\rightarrow$ oscillations of the Gamma, $\varepsilon_1 \simeq 1.9 \times 10^{-2}, \varepsilon_2 \simeq 0.11$
 $m_B = 2 > p \rightarrow$ Delta discontinuous, **Dirac masses in the Gamma**

Proposition 3.1

- constant coefficients, $\Omega = \mathbb{R}$
- $J(\theta) < +\infty$

$$\Phi_\theta^h - u^* \in C^0([0, T]; L^2(\mathbb{R}; \mathbb{R})), \quad \|\Phi_\theta^h - u^*\|_{L^2(Q)}^2 \leq \frac{T^2}{2} e^{2 \max\{\Lambda_{\mathcal{L}^X}, 0\}T} J(\theta)$$

symbol $\lambda_{\mathcal{L}^X, \xi} = \sum_{j=0}^{2p} \gamma_{\mathcal{L}^X, j} (i\xi)^j$ growth bound $\Lambda_{\mathcal{L}^X} = \sup_{\xi \in \mathbb{R}} \operatorname{Re} \lambda_{\mathcal{L}^X, \xi}$

$E_h = +\infty \longrightarrow J(\theta) = +\infty \longrightarrow$ bound **uninformative**

Definition 3.1 – initial-data extension

$$h \in \mathcal{W}, \quad h(x, 0) = g(x), \quad E_h < +\infty$$

Problem

- $\longrightarrow$ construction of initial-data extensions h
- $\longrightarrow$ bounds on E_h

Retained evolution

A retains the term of order $2p$ of $\mathcal{L}^X \rightarrow A$ parabolic

$$\partial_t h = Ah \text{ on } Q, \quad h(x, 0) = g(x), \quad B = \mathcal{L}^X - A \text{ of order } m_B$$

Proposition 4.1

retained part cancelled

$$\mathcal{L}h = \partial_t h - (A + B)h = -Bh, \quad E_h = \|Bh\|_{L^2(Q)}^2$$

Example: Heat Evolution of a Discontinuous Datum

Setting

$$\Omega = \mathbb{R}, \quad p = 1, \quad A = \partial_{xx}^2 \quad \hat{h}(\xi, t) = e^{-t\xi^2} \hat{g}(\xi)$$
$$g = \mathbb{1}_{[-1,1]} \longrightarrow g \in L^2(\mathbb{R}; \mathbb{R}) \setminus H^1(\mathbb{R}) \quad \|\cdot\| = \|\cdot\|_{L^2(\mathbb{R})}$$

$$H^s(\mathbb{R}) = \left\{ f \in L^2(\mathbb{R}) : \frac{\partial^j f}{\partial x^j} \in L^2(\mathbb{R}), 0 \leq j \leq s \right\} = \left\{ f \in L^2(\mathbb{R}) : \int_{\mathbb{R}} (1 + |\xi|)^{2s} |\hat{f}(\xi)|^2 d\xi < +\infty \right\}$$

$s \in \mathbb{N}$, derivatives in the sense of distributions

Integration in time

$$\left\| \frac{\partial^j h}{\partial x^j} \right\|_{L^2(Q)}^2 = \int_{\mathbb{R}} \xi^{2j} \left(\int_0^T e^{-2t\xi^2} dt \right) |\hat{g}(\xi)|^2 d\xi = \int_{\mathbb{R}} \xi^{2j} \frac{1 - e^{-2T\xi^2}}{2\xi^2} |\hat{g}(\xi)|^2 d\xi$$

$$j = 1 \quad \|\partial_x h\|_{L^2(Q)}^2 = \frac{1}{2} \int_{\mathbb{R}} (1 - e^{-2T\xi^2}) |\hat{g}(\xi)|^2 d\xi = \frac{1}{2} (\|g\|^2 - \|h(T)\|^2) < +\infty$$

$$j = 2 \quad \|\partial_{xx}^2 h\|_{L^2(Q)}^2 = \frac{1}{2} \int_{\mathbb{R}} \xi^2 (1 - e^{-2T\xi^2}) |\hat{g}(\xi)|^2 d\xi < +\infty \iff g \in H^1(\mathbb{R}) \text{ here } +\infty$$

Plancherel: $\|g\|^2 = \int_{\mathbb{R}} |\hat{g}|^2$, $\|h(T)\|^2 = \int_{\mathbb{R}} e^{-2T\xi^2} |\hat{g}|^2$ here $|\hat{g}(\xi)|^2 = 2 \sin^2 \xi / (\pi \xi^2)$, so $\int_{\mathbb{R}} \xi^2 |\hat{g}|^2 = +\infty$

time integration $\longrightarrow$ one spatial derivative gained ($p = 1$)

remainder of order $m_B \leq 1 \longrightarrow E_h < +\infty$ $m_B = 2 \longrightarrow E_h < +\infty \iff g \in H^{m_B-p}(\mathbb{R})$

Proposition 4.2

A parabolic, $g \in H^{s_B}(\Omega)$

$$E_h \leq (m_B + 1)C_A \left(\sum_{j=0}^{m_B} \|\gamma_{B,j}\|_{L^\infty(\Omega)}^2 \right) \|g\|_{H^{s_B}(\Omega)}^2$$

$C_A \in (0, +\infty) \rightarrow$ depends on T, p and the coefficients of A only

time integration $\rightarrow$ **gain of p spatial derivatives**

$$s_B = \max(m_B - p, 0)$$

	m_B					
p	0	1	2	3	4	5
1	0	0				
2	0	0	0	1		
3	0	0	0	0	1	2

■ every continuous g

■ $g \in H^1$

■ $g \in H^2$

$m_B \leq 2p - 1$: A retains the order $2p$

Proposition 4.3

constant coefficients, A parabolic

$$(i) \quad m_B > p \implies (E_h < +\infty \iff g \in H^{s_B}(\Omega))$$

$$(ii) \quad m_B \leq p \iff (E_h < +\infty \quad \forall g \in C^0(\Omega; \mathbb{R}) \cap L^2(\Omega; \mathbb{R})) \iff \gamma_{A,j} = \gamma_{\mathcal{L}^x,j} \quad \forall j > p$$

Measured (3 seeds, fixed protocol)

- three test cases $\rightarrow m_B \leq p$: **smallest median error** among the compared choices, control excluded
- order four, median error
 - $A = 0$ (Lagaris) $\rightarrow 0.151$
 - Chen–Mangasarian (Zhang et al.) $\rightarrow 33.9$
 - A : order 4 $\rightarrow 4.90 \times 10^{-4}$
- butterfly, Greeks at $t = T$, $m_B = 2 > p \rightarrow$ Dirac masses in the Gamma

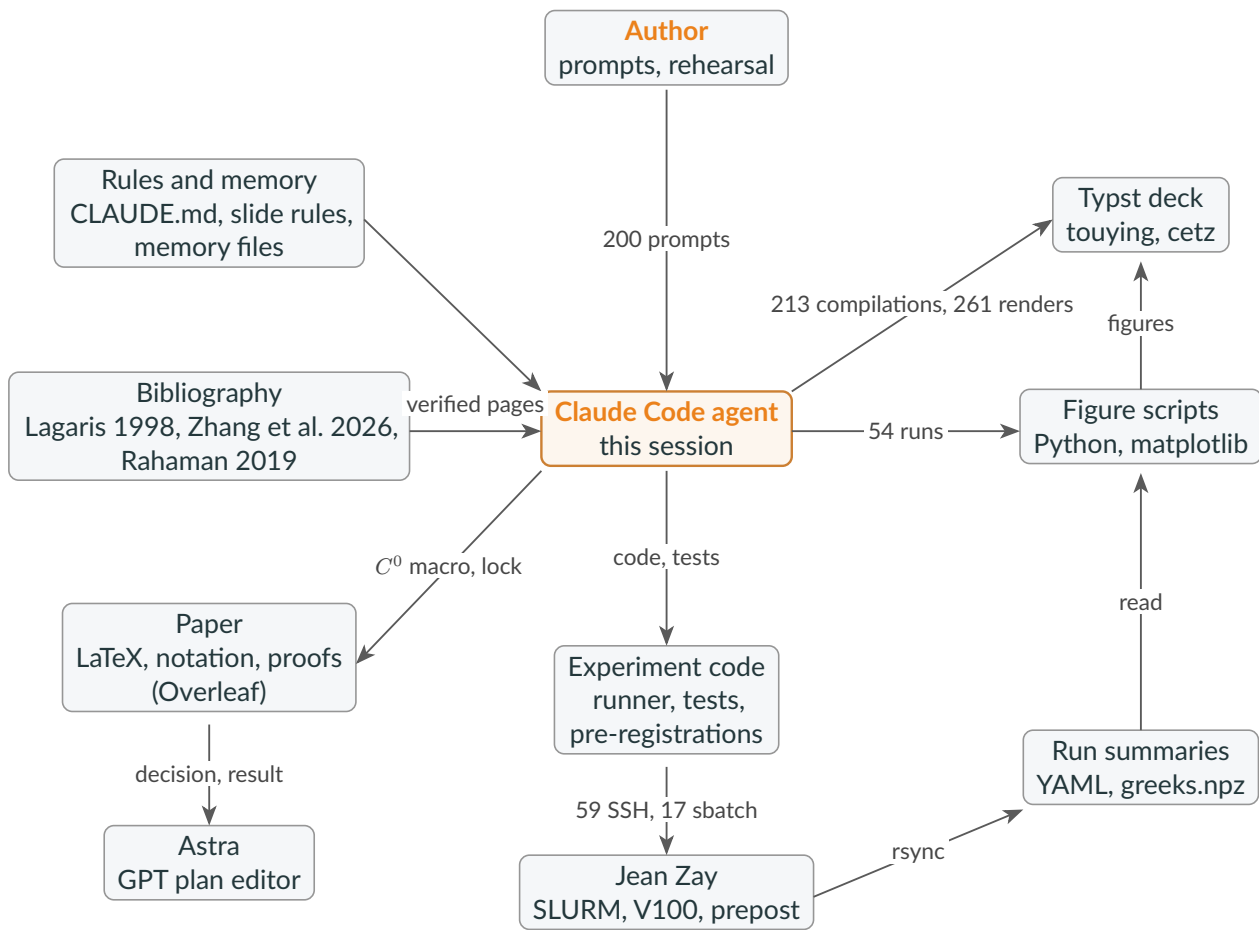
Proved (A parabolic)

- $h = e^{tA}g \rightarrow$ exact initial trace, forcing $-Bh$ (Proposition 4.1)
- E_h bounded by $\|g\|_{H^{s_B}(\Omega)}^2$, $s_B = \max(m_B - p, 0)$ (Proposition 4.2)
- $E_h < +\infty$ for every $g \in C^0(\Omega; \mathbb{R}) \cap L^2(\Omega; \mathbb{R}) \iff m_B \leq p$ (Proposition 4.3)

Open question

$\rightarrow$ relation between forcing energy and trained error

Appendix: How These Slides Were Made



Session statistics (10–14 October 2026, measured on the transcript)

Quantity	Count
Author prompts	200
in French	126
on the slides ^a	107
with a pasted slide excerpt	37
Words per prompt (median)	14
Typst compilations	213
Page renders inspected	261
Scripted edits of the deck	198
Figure regenerations	54
Git commits (all repositories)	53
Effective time on the slides ^b	≈ 7.5 h
Effective time, whole session ^b	≈ 14.2 h

^a keyword classification (slide, figure, caption, pasted excerpt, ...), approximate; words counted without pasted excerpts; the other prompts concern the experiments, the paper and scientific questions

^b sum of the intervals between transcript events shorter than 15 min, assigned to the topic of the current prompt; 6.4 to 9.3 h on the slides for thresholds of 10 to 30 min; wall-clock span 87.5 h

First edition of these slides, within a first appropriation and assimilation of this work