Friedrich-Alexander-Universität
Research Center for
Mathematics of Data | MoD

Qualitative numerics

Discrete versus continuous mechanics

Enrique Zuazua[1]

FAU Erlangen-Nürnberg
Alexander von Humboldt Professorship
ERC AdvG CoDeFeL

September 22, 2024

Unterstützt von / Supported by
Alexander von Humboldt
Stiftung/Foundation

European Research Council
Established by the European Commission

[1] Based mainly on joint work with Liviu Ignat (Bucharest)

Presentation Outline

The problem

In the linear context (spectral theory) best constants are characterised by a Rayleigh quotient

$$C_{A,B,X} = \inf_{u \in X} \frac{a(u,u)}{b(u.u)}.$$

Other, non-Hilbertian constants, such as the Sobolev one, can rather be written as

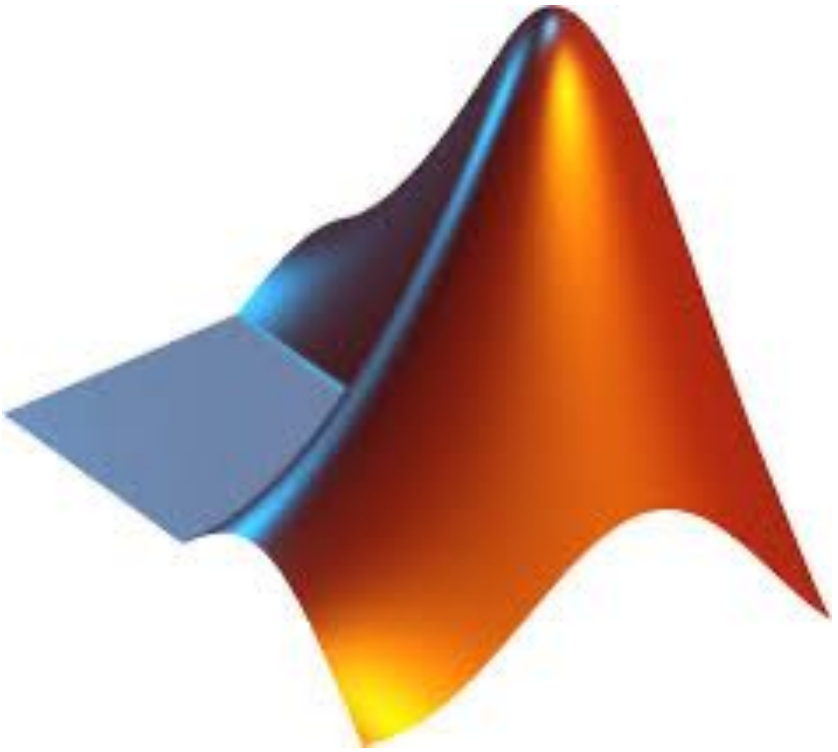
$$C_{A,B,X} = \inf_{u \in X} \frac{\|u\|_A}{\|u\|_B}.$$

From a numerical analysis perspective the most natural and elementary question is: How is $C_{A,B,X}$ approximated by $C^h_{A_h,B_h,X_h}$, when h stands for a numerical approximation procedure?

This is one of the most elementary problems in qualitative numerics. More complex ones: Long time asympotics, dispersion, inverse problems, controls,....

Poincaré constant

The Poincaré constant:



$$\lambda_1(\Omega) = \min_{u \in H_0^1(\Omega)} \frac{\int_\Omega |\nabla u|^2 dx}{\int_\Omega u^2 dx}$$

$$\begin{cases} -\Delta u = \lambda u & \text{in } \Omega \\ u = 0, & \text{on } \partial\Omega \end{cases}$$

FEM approximation literature

[1] E. Ernst and P. Le Tallec (2004), "Numerical Approximation of Poincar and Friedrichs Constants." This paper is often cited for its comprehensive numerical methods to compute these constants using finite element approximations.

[2] I. Babuška, J.E. Osborn, Eigenvalue problems. In Handbook of Numerical Analysis, Vol. II, North-Holland, Amsterdam, pp. 641-787 (1991).

[3] D. Boffi, Finite element approximation of eigenvalue problems, Acta Numerica, pp. 1-120 (2010).

[4] Notes/slides by Jaap van der Vegt, Daniele Boffi, etc...

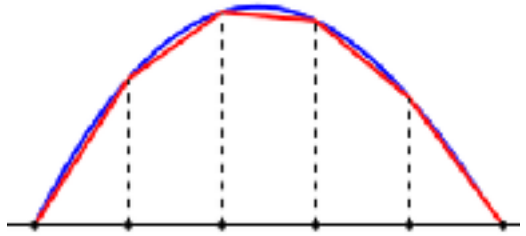
P_1 FEM

- V_h, the P_1-FEM approximating subspace, being included in V:

$$\lambda_1 = \min_V \frac{\|u\|_A^2}{\|u\|_B^2} \leq \min_{V_h} \frac{\|u\|_A^2}{\|u\|_B^2} = \lambda_{1h}$$

- In the P_1 context most references indicate (Raviart-Thomas, Brenner-Scott, etc...)

$$\lambda_1 \leq \lambda_{1h} \leq \lambda_1 + Ch^2$$



In fact

$$\lambda_1 < \lambda_1 + ch^2 \leq \lambda_{1h} \leq \lambda + Ch^2$$

Can one aim a more explicit asymptotic expansion $\lambda_{1h} \sim \lambda_1 + c^* h^2$? Of course in $1-d$ and on uniform meshes this can be done, since computations are explicit.

Sharp lower bound $\lambda_1 + ch^2 \leq \lambda_{1h}$

[1] I. Babuška and J. Osborn. Eigenvalue problems. In *Handbook of numerical analysis, Vol. II*, volume II of *Handb. Numer. Anal.*, pages 641–787. North-Holland, Amsterdam, 1991. Section 8, p. 700
[2] F. Chatelin. *Spectral approximation of linear operators*, volume 65 of *Classics in Applied Mathematics*. SIAM 2011. Reprint of the 1983 original Prop. 6.30, p. 315.

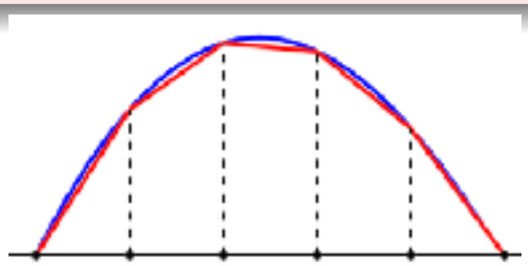
For general operators and general approximation spaces V_h of the Hilbert space V

$$\lambda_{1h}(V_h) - \lambda_1(V) \simeq d_V(u_1, V_h)^2$$

where u_1 is the first eigenvalue of the continuous problem

Task: Compute

$$d_V(u_1, V_h)^2 = \min_{v \in V_h} \int_\Omega |\nabla u_1 - \nabla v_h|^2 dx?$$



Computation of the distance

Classical error estimates apply ("worst case scenario")

$$\inf_{v \in V_h} \|u - v_h\|_V \leq h\|u\|_{H^2(\Omega)}$$

We need a lower bound

$$d_V(u_1, V_h) \gtrsim h$$

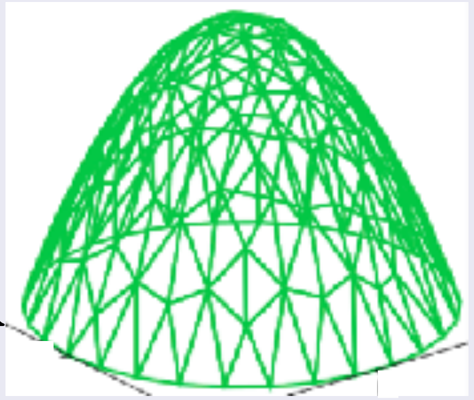
This is essentially due to the convexity properties of the first eigenfunction

$$|u_1''| = \lambda_1 |u_1| \geq c > 0 \quad \text{away from the boundary.}$$

Lemma

For any $-\infty < a < b < \infty$ *and* $u \in C^2((a,b))$

$$\Longrightarrow \int_a^b |u'(r) - A|^2 dr \geq \frac{(b-a)^3}{12} \inf_{r \in [a,b]} |u''(r)|^2, \ \forall A \in \mathbb{R}$$



Similar lower bounds can be obtained in higher dimensions working on a neighbourhood of the max of the eigenfunction.

Sobolev constant

$$2^* = 2N/(N-2),\, N \geq 3.$$

$$S(N) = \inf_{u \in H^1(\mathbb{R}^N)} \frac{\|\nabla u\|_{L^2(\mathbb{R}^N)}}{\|u\|_{L^{2^*}(\mathbb{R}^N)}}. \tag{1}$$

Best Constant in Sobolev Inequality (*).

GIORGIO TALENTI (Firenze)

- inf = min
- $S(N)$ and $\mathcal{M}$(=the set of minimisers) are explicit (G. Talenti, 1975)

$$U_{a,b,x_0}(x) = \frac{a}{(1+b|x-x_0|^2)^{\frac{N-2}{2}}},\; a \in \mathbb{R}\setminus\{0\},\; b>0,\, x_0 \in \mathbb{R}^N$$

$$-\Delta U = S(N)\|U\|^{2-2^*}_{L^{p^*}(\mathbb{R}^N)}|U|^{2^*-2}U,\; \text{in } \mathbb{R}^N.$$

- $\mathcal{M} = N+2$ manifold in $H^1(\mathbb{R}^N)$
- When Ω bounded domain, $S(N,\Omega) = S(N)$ but inf in $H^1_0(\Omega)$ is not attained

L. Ignat & E. Z., 2024

Theorem (Sobolev constants)

Let $\Omega \subset \mathbb{R}^N$ be bounded, $N \geq 3$, and V_h the space of P_1 finite elements space on Ω. Then the FEM approximation of $S(N)$, $S_h(N)$, satisfies

$$S_h(N) - S(N) \sim h^{\frac{2(N-2)}{N}}.$$

This improves the rate $h^{1/3}$ for $N = 3$ in [P. F. Antonietti and A. Pratelli. Finite element approximation of the Sobolev constant. Numer. Math. 2011.]

The Sobolev deficit

$$\delta(u) = \frac{\|Du\|_{L^p(\mathbb{R}^N)}}{\|u\|_{L^{p*}(\mathbb{R}^N)}} - S_{p,N}, \ \forall\ u \in \dot{W}^{1,p}(\mathbb{R}^N).$$

Bianchi and Egnel [JFA91] (p=2) and Figally and Zgang [Duke22] ($1 < p < N$)

$$\delta(u) \geq c \inf_{v \in \mathcal{M}} \left(\frac{\|Du - Dv\|_{L^p(\mathbb{R}^N)}}{\|Du\|_{L^p(\mathbb{R}^N)}}\right)^{\max\{2,p\}} \tag{2}$$

Lemma

There is a constant $C = C(p, N)$ and a positive number ε_0 such that

$$\delta(u) \leq C \inf_{v \in \mathcal{M}} \left(\frac{\|Du - Dv\|_{L^p(\mathbb{R}^N)}}{\|Du\|_{L^p(\mathbb{R}^N)}}\right)^{\min\{2,p\}} \tag{3}$$

holds for all $u \in W^{1,p}(\mathbb{R}^N)$ satisfying

$$d(u, \mathcal{M}) = \inf_{v \in \mathcal{M}} \|Du - Dv\|_{L^p(\mathbb{R}^N)} \leq \varepsilon_0 \|Du\|_{L^p(\mathbb{R}^N)}.$$

Sketch of proof

Let $w_h \in V_h$ with $\|Dw_h\|^2_{L^2(\mathbb{R}^N)} = 1$ be a discrete FEM minimiser:

$$S_h := \frac{\|Dw_h\|_{L^2(\mathbb{R}^N)}}{\|w_h\|_{L^{2^*}(\mathbb{R}^N)}}$$

Using estimates for the deficit

$$\inf_{U\in\mathcal{M}} d_V^2(U, w_h) \lesssim S_h - S \lesssim \inf_{U\in\mathcal{M}} d_V^2(U, V_h).$$

- One still needs to prove that

$$h^{2/3} \lesssim \inf_{U\in\mathcal{M}} d_V^2(U, w_h), \qquad \inf_{U\in\mathcal{M}} d_V^2(U, V_h) \lesssim h^{2/3}$$

Each of them is tricky :)

Upper/lower bounds

- **Upper bound**

Take $u = U_{1,\lambda,0}$ and $v_h = I_h(u - u_{|\partial B_h})$ extended with zero outside B_h

$$\int_{\mathbb{R}^N} |DU_\lambda - Dv_h|^2 = \int_{B_h^c} |DU_\lambda|^2 + \int_{B_h} |DU_\lambda - Dv_h|^2 \lesssim \lambda^{-2} + (h\lambda^2)^2$$

and optimize in λ, i,e. $\lambda = h^{-1/3}$.

- **Lower bound**

Take where $a_h > 0$, $\lambda_h > 0$ and $x_h \in \mathbb{R}^N$ $(\dim(\mathcal{M}) = N + 2)$

$$d^2(U_{a_h,\lambda_h,x_h}, w_h) = \inf_{U \in \mathcal{M}} d_V^2(U, w_h) \lesssim S_h - S \lesssim \inf_{U \in \mathcal{M}} d_V^2(U, V_h) \lesssim h^{2/3}$$

Few steps: $a_h > 1/2$, $x_h \in B_h$, $(\lambda_h^2 h)_{h>0}$ bounded, otherwise the above upper bound is false. Under these assumption

$$\int_{\mathbb{R}^N} |DU_{a_h,\lambda_h,x_h} - Dv_h|^2 \gtrsim \lambda_h^{-2} + h^2\lambda_h^4 \gtrsim h^{2/3}.$$

Hardy's inequality

$$H^*(\Omega) = \inf_{u \in H_0^1(\Omega)} \frac{\|\nabla u\|^2_{L^2(\Omega)}}{\|\frac{u}{|x|}\|^2_{L^2(\Omega)}}, \quad 0 \in \Omega \tag{5}$$

- $H^*(\Omega) = H^*(\mathbb{R}^N) = \frac{(N-2)^2}{4}$, $N \neq 2$
- H^* is not attained in $H_0^1(\Omega)$. Any possible minimizer should be of the form

$$u(r) = r^{-\frac{N-2}{2}}(a_1 + a_2 \log(r)) \notin W^{1,2}_{loc}(\mathbb{R}^N)$$

but it belongs to a bigger space $\mathcal{H}(\Omega)$, the completion of $C_c^\infty(\Omega)$ with respect to the norm [2]

$$\|u\|^2_H = \int_\Omega |\nabla u|^2 - H^* \frac{u^2}{|x|^2}, \tag{6}$$

- $\mathcal{H}$ is isometric with the space $W_0^{1,2}(|x|^{-(N-2)}dx, \Omega)$

$$u = Tv = |x|^{-(N-2)/2} v.$$

---

[2] J. L. Vázquez, J. L., E. Z. (2000). The Hardy inequality and the asymptotic behaviour of the heat equation with an inverse-square potential. Journal of Functional Analysis, 173(1), 103-153.

Theorem

Let us consider $\Omega \in \mathbb{R}^N$, $N \geq 3$, and V_h the space of linear finite elements space on Ω. Then the corresponding discrete minimizers satisfy

$$H^* - H_h^* \simeq \frac{1}{|\log h|^2}.$$

Sketch of proof

• For the lower bound: Hardy inequality with **logarithmic reminder**. [3]

• For the upper bound: take a **pseudo-minimizer** $u(x) = |x|^{-\frac{N-2}{2}} \log^{1/2}(\frac{1}{|x|})$, regularize it, and approximate the regularization with P_1 functions.

[3] I. Peral, F. Soria, Elliptic and Parabolic Equations involving the Hardy-Leray Potential, 2021.

Presentation Outline

Numerical hypocoercivity for the Kolmogorov equation

A. Porretta & E. Zuazua. (2017). Numerical hypocoercivity for the Kolmogorov equation, Mathematics of Computation.

The Kolmogorov equation (1931)

$$\partial_t f - \partial_{xx} f - x\partial_y f = 0$$

can be rewritten:

$$f_t + AA^* f + Bf = 0$$

where $A = \partial_x$ and $B = x\partial_y$.
The energy identity shows an apparent insufficient dissipation rate:

$$\frac{1}{2}\frac{d}{dt}\|f(t)\|_{L^2}^2 = -\|\partial_x f(t)\|_{L^2}^2.$$

But Kolmogorov found an explicit representation of the fundamental solution exhibiting a gaussian behavior both in (x, y). In particular

$$\|f(t)\|_{L^2} + \sqrt{t}\|\partial_x f(t)\|_{L^2} + t^{\frac{3}{2}}\|\partial_y f(t)\|_{L^2} \leq C\|f_0\|_{L^2}.$$

The reason for the gain of decay on $\partial_y f$ is $[\partial_x, x\partial_y]f = \partial_y f$.
L. Hörmander (1967) coined the term **hypoellipticity**.

To develop numerical schemes preserving the decay as $t \to \infty$, we adopt the view point of **hypocoercivity** by Villani, Hérau, *et al.*, which consists in considering a Lyapunov functional of the form

$$\mathcal{L}(t) = \lambda \|f(t)\|_{L^2}^2 + at\|\partial_x f(t)\|_{L^2}^2 + bt^2(\partial_x f, [\partial_x, x\partial_y]f(t)) + ct^3\|\partial_y f\|_{L^2}^2.$$

Adjusting the coefficients $a, b, c > 0$, we get

$$\frac{d}{dt}(\mathcal{L}(t)) < 0$$

exhibiting sharp decay rates.

A similar argument allows to prove the uniform decay for finite-difference numerical approximation schemes of the form

$$\frac{d}{dt} f_{i,j}(t) - \frac{1}{h^2}\left[f_{i+1,j}(t) + f_{i-1,j}(t) - 2f_{i,j}(t))\right] - x_i \frac{[f_{i,j+1}(t) - f_{i,j-1}(t)]}{2h} = 0.$$

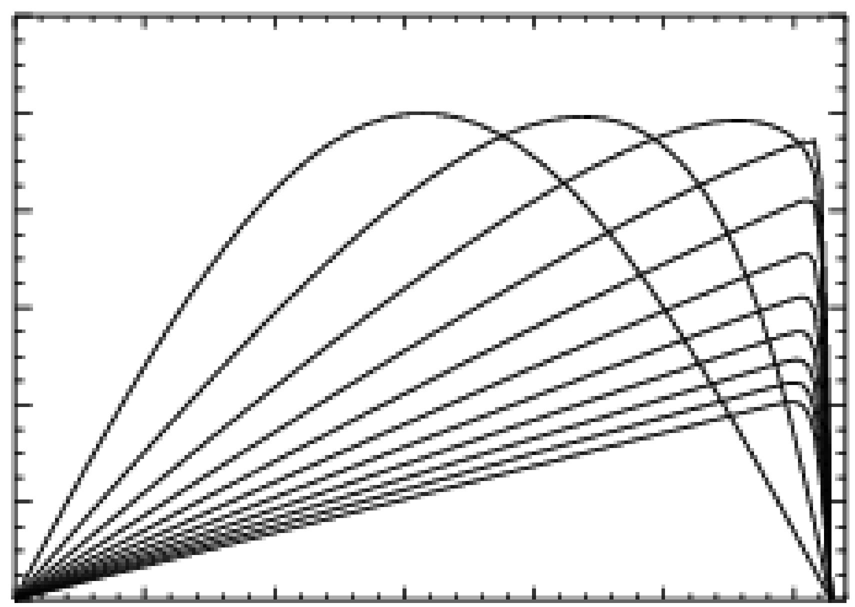
To be compared with the more subtle phenomena occurring in the context of waves and dispersive equations, due to high frequency spurious numerical solutions.

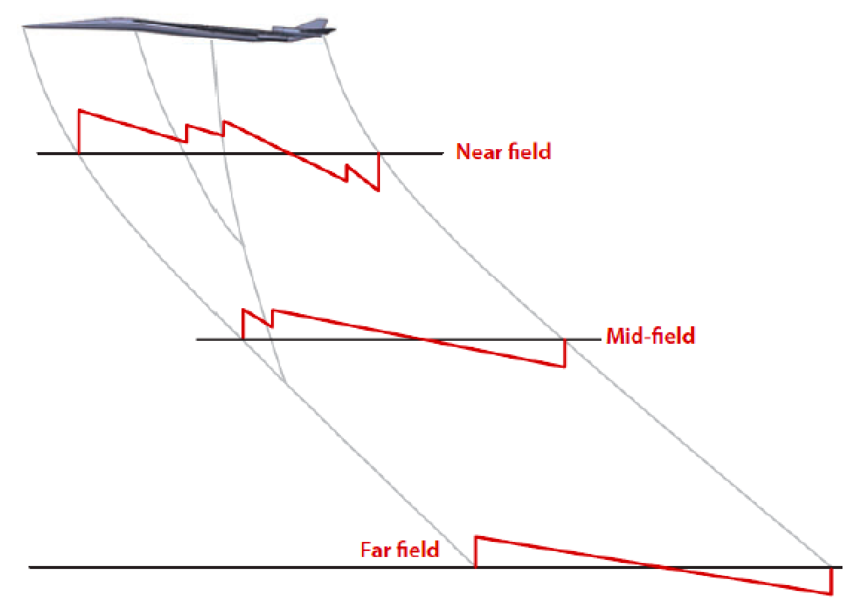
Presentation Outline

Time irreversibility by nonlinearity

For hyperbolic Burgers equation (and more generally for first order non-linear conservation laws) irreversibility occurs because of shock formation:[4]

$$u_t + \partial_x[u^2] = 0.$$





[4] Juan J. Alonso and Michael R. Colonno, Multidisciplinary Optimization with Applications to Sonic-Boom Minimization, Annu. Rev. Fluid Mech. 2012, 44:505 – 26.

Conservative schemes

Let us consider now numerical approximation schemes for the inviscid problem :

$$\begin{cases} u_j^{n+1} = u_n^j - \dfrac{\Delta t}{\Delta x}\left(g_{j+1/2}^n - g_{j-1/2}^n\right), & j \in \mathbb{Z}, n > 0. \\ u_j^0 = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} u_0(x)dx, & j \in \mathbb{Z}. \end{cases}$$

The approximated solution u_Δ is given by

$$u_\Delta(t,x) = u_j^n, \quad x_{j-1/2} < x < x_{j+1/2}, \, t_n \le t < t_{n+1},$$

where $t_n = n\Delta t$ and $x_{j+1/2} = (j + \frac{1}{2})\Delta x$.

Is the large tine dynamics of these discrete systems, a discrete version of the continuous one?

3-point conservative schemes

1. Lax-Friedrichs

$$g^{LF}(u,v) = \frac{u^2+v^2}{4} - \frac{\Delta x}{\Delta t}\left(\frac{v-u}{2}\right),$$

2. Engquist-Osher

$$g^{EO}(u,v) = \frac{u(u+|u|)}{4} + \frac{v(v-|v|)}{4},$$

3. Godunov

$$g^{G}(u,v) = \begin{cases} \min_{w\in[u,v]} \frac{w^2}{2}, & \text{if } u \le v, \\ \max_{w\in[v,u]} \frac{w^2}{2}, & \text{if } v \le u. \end{cases}$$

Properties

These three schemes are well-known to satisfy the following properties:

- They converge to the entropy solution
- They are monotonic
- They preserve the total mass of solutions
- They are OSLC consistent:

$$\frac{u_{j-1}^n - u_{j+1}^n}{2\Delta x} \leq \frac{2}{n\Delta t}$$

- $L^1 \to L^\infty$ decay with a rate $O(t^{-1/2})$
- Similarly they verify uniform BV_{loc} estimates

Asymptotic correctness as $t \to \infty$?

- All these methods converge in the classical sense of numerical analysis.
- This refers to convergence in finite time intervals $[0, T]$!!!
- But do they behave correctly as $t \to \infty$?
- Note that, computationally, roughly, you can choose Δx and Δt, but once you do this, you have to rely on what simulations give as $t \to \infty$.

  And this is relevant when solving numerically the minimisation problem since a gradient descent method requires the iterative resolution of the forward and adjoint dynamics many times, producing then the effect of solving the PDE over long time intervals.

Numerical viscosity

We can rewrite three-point monotone schemes in the form

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{(u_{j+1}^n)^2 - (u_{j-1}^n)^2}{4\Delta x} = R(u_j^n, u_{j+1}^n) - R(u_{j-1}^n, u_j^n)$$

where the numerical viscosity R can be defined in a unique manner as

$$R(u,v) = \frac{Q(u,v)(v-u)}{2} = \frac{\lambda}{2}\Big(\frac{u^2}{2} + \frac{v^2}{2} - 2g(u,v)\Big).$$

For instance:

$$R^{LF}(u,v) = \frac{v-u}{2\Delta t},$$

$$R^{EO}(u,v) = \frac{\lambda}{4}(v|v| - u|u|),$$

$$R^{G}(u,v) = \begin{cases} \frac{\lambda}{4}\operatorname{sign}(|u| - |v|)(v^2 - u^2), & v \le 0 \le u, \\ \frac{\lambda}{4}(v|v| - u|u|), & \text{elsewhere.} \end{cases}$$

Linear versus nonlinear numerical viscosity

The main difference between LF and EO/G schemes is that in the first one viscosity is linear and in the later ones it is nonlinear (quadratic).

What is best linear or nonlinear?
Nonlinear models are more adaptive but also more unpredictable

Acta Numerica (1998), pp. 51–150

Nonlinear approximation

Ronald A. DeVore*
Department of Mathematics,
University of South Carolina,
Columbia, SC 29208, USA
E-mail: devore@math.sc.edu

Lax-Friedrichs = Viscous

Theorem (Lax-Friedrichs scheme)

Consider $u_0 \in L^1(\mathbb{R})$ and Δx and Δt such that $\lambda \|u^n\|_{\infty,\Delta} \leq 1$, $\lambda = \Delta t/\Delta x$. Then, for any $p \in [1,\infty)$, the numerical solution u_Δ given by the Lax-Friedrichs scheme satisfies

$$\lim_{t\to\infty} t^{\frac{1}{2}(1-\frac{1}{p})} \|u_\Delta(t) - w(t)\|_{L^p()} = 0,$$

where the profile $w = w_{M_\Delta}$ is the unique solution of

$$\begin{cases} w_t + \left(\frac{w^2}{2}\right)_x = \frac{(\Delta x)^2}{2\Delta t} w_{xx}, & x \in \mathbb{R}, t > 0, \\ w(0) = M_\Delta \delta_0, \end{cases}$$

with $M_\Delta = \int u^0_\Delta$.

[5]

[5] L. Ignat, A. Pozo & E. Z., Large-time asymptotics, vanishing viscosity and numerics for 1-D scalar conservation laws, Math of Computation, 84(294), 2015.

Engquist-Osher and Godunov = Inviscid

Theorem (Engquist-Osher and Godunov schemes)

Consider $u_0 \in L^1(\mathbb{R})$ and Δx and Δt such that $\lambda \|u^n\|_{\infty,\Delta} \leq 1$, $\lambda = \Delta t / \Delta x$. Then, for any $p \in [1, \infty)$, the numerical solutions u_Δ given by Engquist-Osher and Godunov schemes satisfy the same asymptotic behavior but for the hyperbolic $N - wave$ $w = w_{p_\Delta, q_\Delta}$ unique solution of

$$\begin{cases} w_t + \left(\frac{w^2}{2}\right)_x = 0, & x \in \mathbb{R}, t > 0, \\ w(0) = M_\Delta \delta_0, & \lim_{t \to 0} \int_0^x w(t,z)dz = \begin{cases} 0, & x < 0, \\ -p_\Delta, & x = 0, \\ q_\Delta - p_\Delta, & x > 0, \end{cases} \\ \end{cases}$$

with $M_\Delta = \int u_\Delta^0$ and $p_\Delta = -\min_{x\in} \int_{-\infty}^x u_\Delta^0(z)dz$ and $q_\Delta = \max_{x\in} \int_x^\infty u_\Delta^0(z)dz$.

Why?

- For the Lax-Friedrichs scheme diffusion is linear, it is invariant with respect to time. it remains asymptotically, leading to the viscous effect

$$-\frac{(\Delta x)^2}{2\Delta t} w_{xx}$$

as $t \to \infty$.

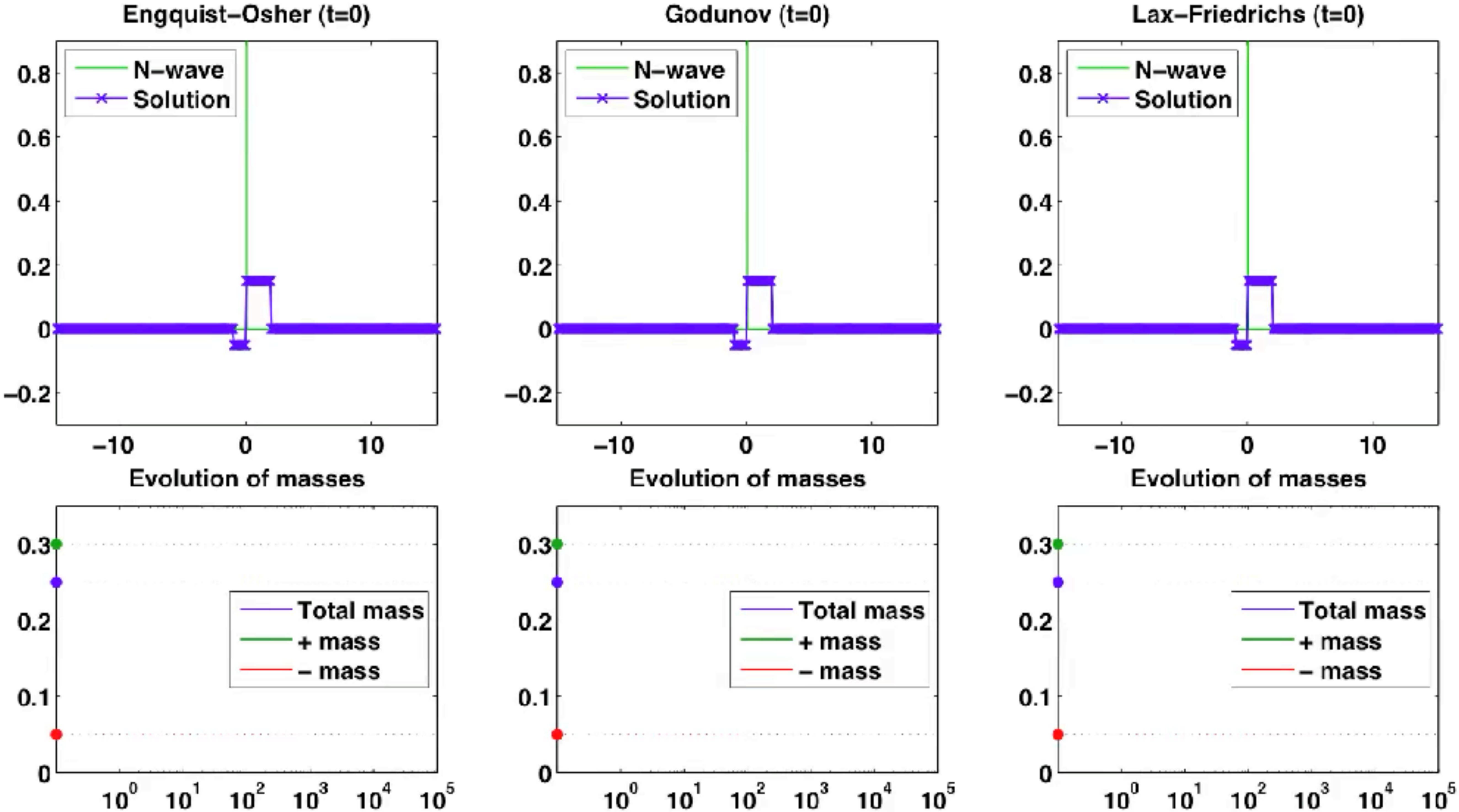
- But, for the Engquist-Osher and Godunov schemes the viscosity is non-linear of the order, roughly, of

$$-(w^2)_{xx} = -(w \cdot w)_{xx}$$

. Taking into account that we have the uniform a priori bound

$$|w(x,t)| \leq Ct^{-1/2}$$

they amount of viscosity decreases as $t \to \infty$ and eventually disappears.



Engquist–Osher (t=0)
Godunov (t=0)
Lax–Friedrichs (t=0)
N–wave
Solution
Evolution of masses
Total mass
+ mass
– mass

Presentation Outline

Observing waves

The control problem is equivalent to the observability one for the adjoint wave equation:

$$\begin{cases} \varphi_{tt} - \varphi_{xx} = 0, & 0 < x < 1, 0 < t < T \\ \varphi(0,t) = \varphi(1,t) = 0, & 0 < t < T. \end{cases}$$

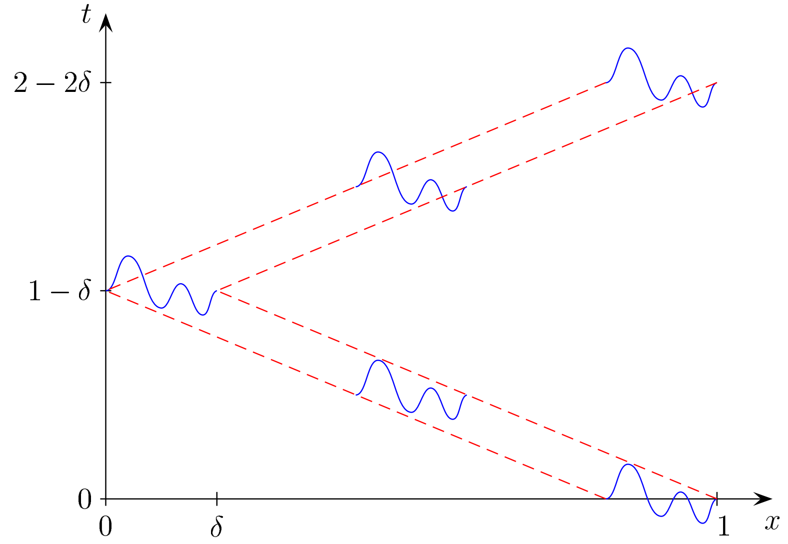
Namely,

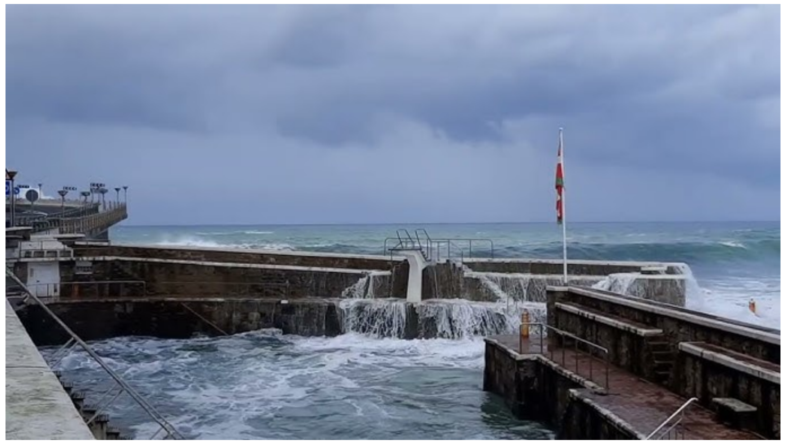
$$E(0) \le C(T) \int_0^T |\varphi_x(1,t)|^2 \, dt.$$

for the conserved energy

$$E(t) = \frac{1}{2} \int_0^1 \left[|\varphi_x(x,t)|^2 + |\varphi_t(x,t)|^2 \right] dx = E(0), \quad \forall 0 \le t \le T.$$

The observability inequality holds if and only if $T \ge 2$.





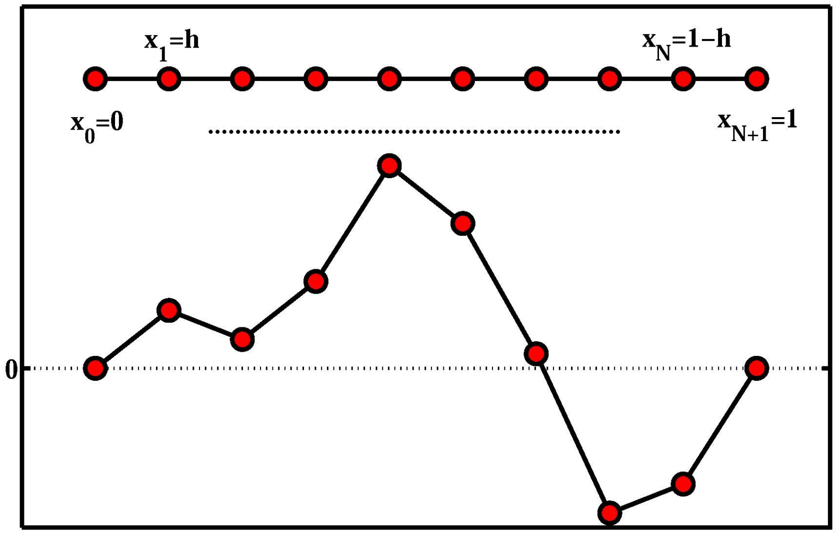
The discrete approach: Discretize and then control

Set $h = 1/(N+1) > 0$ and consider the mesh

$$x_0 = 0 < x_1 < ... < x_j = jh < x_N = 1 - h < x_{N+1} = 1,$$

and the finite difference semi-discrete approximation

$$\begin{cases} \varphi_j'' - \frac{1}{h^2}\left[\varphi_{j+1} + \varphi_{j-1} - 2\varphi_j\right] = 0, & 0 < t < T, \, j = 1, \ldots, N \\ \varphi_j(t) = 0, & j = 0, \, N+1, \, 0 < t < T \\ \varphi_j(0) = \varphi_j^0, \, \varphi_j'(0) = \varphi_j^1, & j = 1, \ldots, N. \end{cases}$$



From finite-dimensional dynamical systems to infinite-dimensional ones in purely conservative dynamics.....

WHY?

Discrete solution:

$$\vec{\varphi} = \sum_{k=1}^{N} \left(a_k \cos\left(\sqrt{\lambda_k^h} t\right) + \frac{b_k}{\sqrt{\lambda_k^h}} \sin\left(\sqrt{\lambda_k^h} t\right) \right) \vec{w}_k^h.$$

Continuous solution:

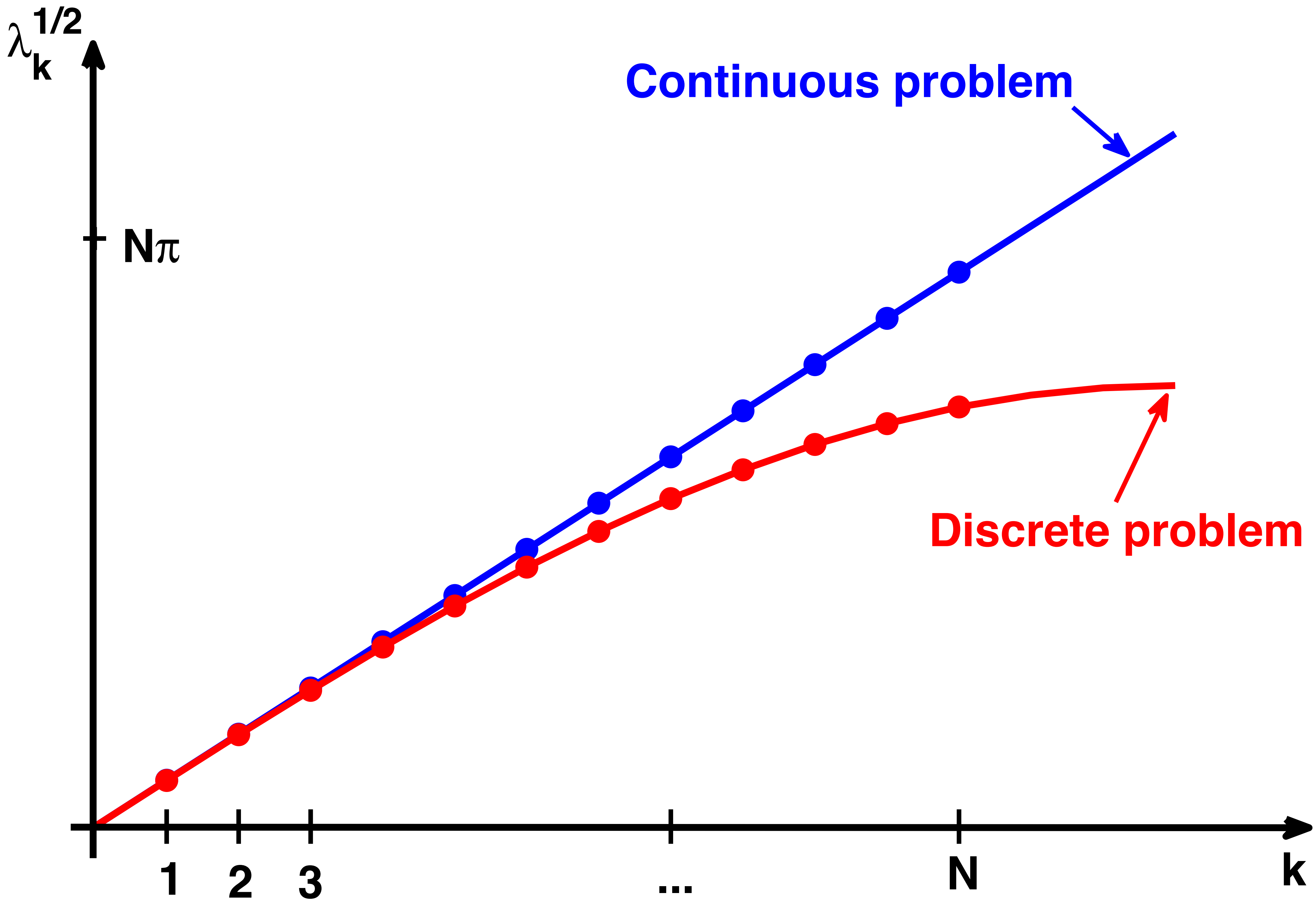
$$\varphi = \sum_{k=1}^{\infty} \left(a_k \cos(k\pi t) + \frac{b_k}{k\pi} \sin(k\pi t) \right) \sin(k\pi x)$$

And spectral convergence holds:

$$\lambda_k^h = \frac{4}{h^2} \sin^2\left(\frac{k\pi h}{2}\right)$$

$$\lambda_k^h \to \lambda_k = k^2\pi^2, \text{ as } h \to 0$$

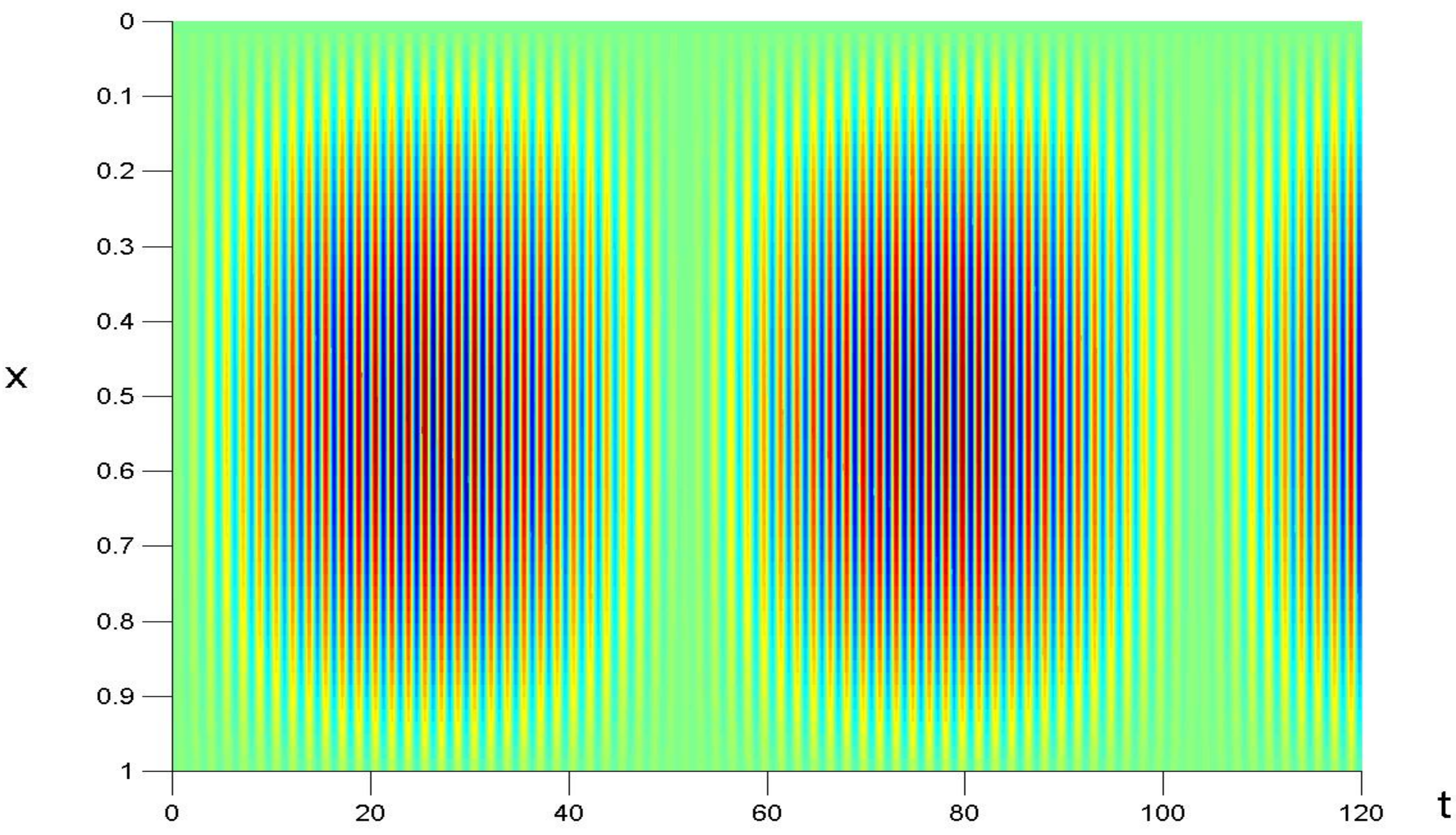
$$w_k^h = (w_{k,1}, \ldots, w_{k,N})^T : w_{k,j} = \sin(k\pi jh),\ k, j = 1, \ldots, N.$$


$\lambda_k^{1/2}$
Continuous problem
$N\pi$
Discrete problem
1
2
3
...
N
k

A numerical ghost

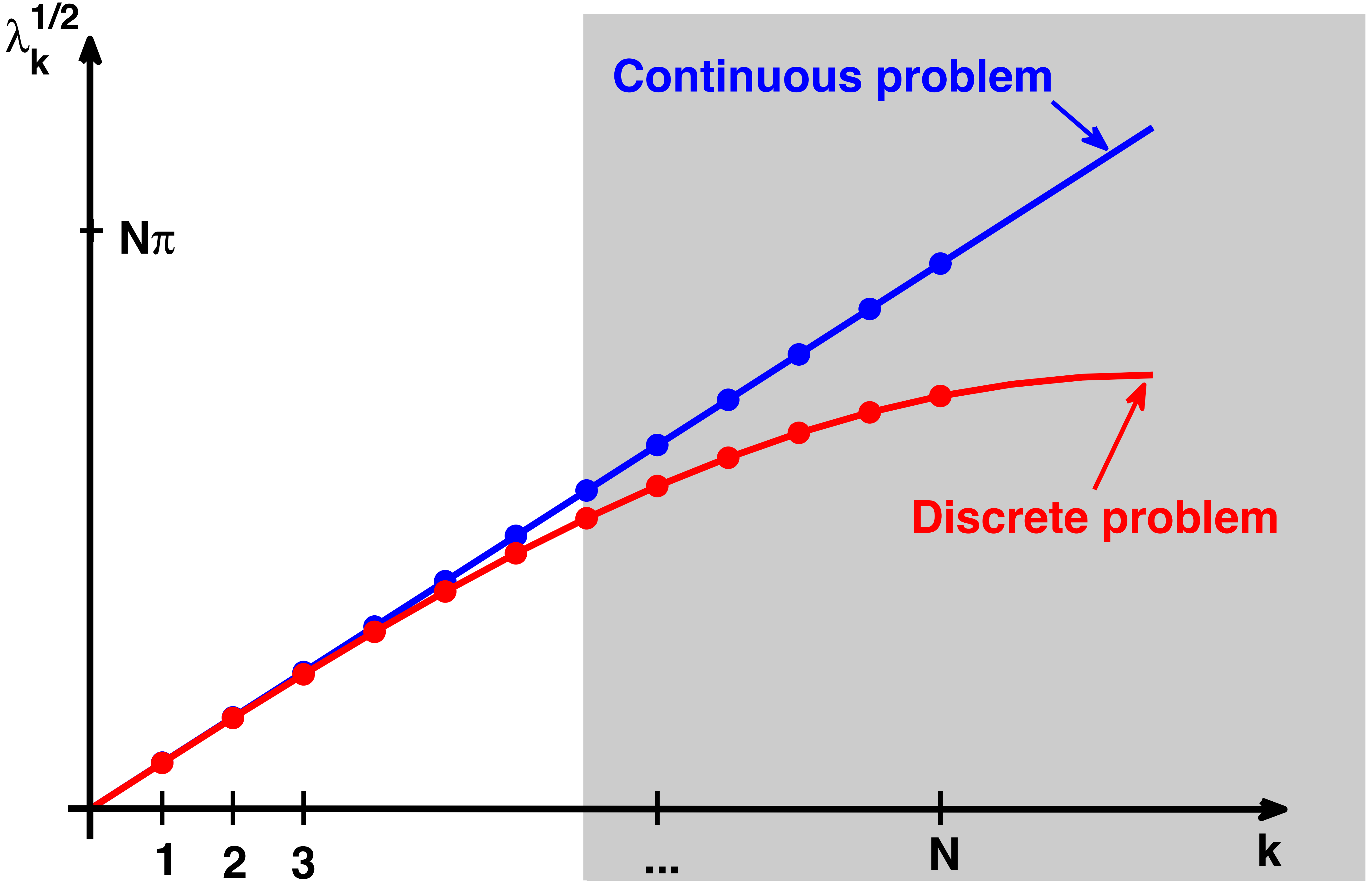
$$\vec{\varphi} = \exp\left(i\sqrt{\lambda_N(h)}\,t\right)\vec{w}_N - \exp\left(i\sqrt{\lambda_{N-1}(h)}\,t\right)\vec{w}_{N-1}.$$

Spurious semi-discrete wave combining the last two eigenfrequencies with very little gap: $\sqrt{\lambda_N(h)} - \sqrt{\lambda_{N-1}(h)} \sim h$.



$h = 1/61$, $(N = 60)$, $0 \le t \le 120$.

A remedy: Fourier filtering

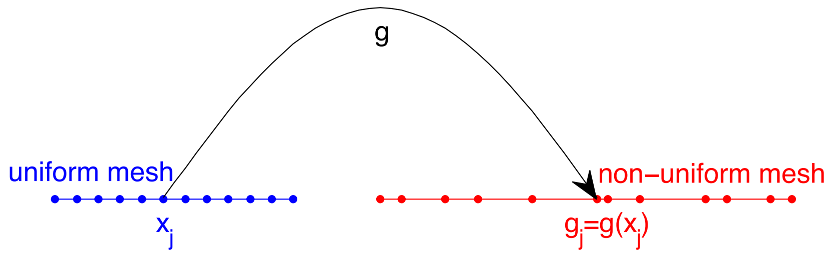


Numerics on non-uniform meshes

(1) A. Marica & E. Z. FoCM, (2015) 15:1571–1636.
(2) A. Ervedoza, A. Marica & E. Z. Numerical meshes ensuring uniform observability of 1d waves: construction and analysis, IMA J. Numer. Anal. 36 (2016), no. 2, 503542.

$$\rho(x)u_{tt} - (\sigma(x)u_x)_x = 0$$

$$\mathcal{E}_{\rho,\sigma}(u^0,u^1) := \frac{1}{2}\int_{\mathbb{R}} (\rho(x)|u_t(x,t)|^2 + \sigma(x)|u_x(x,t)|^2)\,dx$$



$$\rho(g_j)u_{j,tt}(t) - \frac{\sigma(g_{j+1/2})\frac{u_{j+1}(t)-u_j(t)}{g_{j+1}-g_j} - \sigma(g_{j-1/2})\frac{u_j(t)-u_{j-1}(t)}{g_j-g_{j-1}}}{\frac{g_{j+1}-g_{j-1}}{2}} = 0. \qquad (7)$$

$$\mathcal{E}^h_{\rho,\sigma,g}(\mathbf{u}^{h,0},\mathbf{u}^{h,1}) := \frac{h}{2}\sum_{j\in}[\partial^h g_j\rho(g_j)|u_{j,t}(t)|^2 + \frac{\sigma(g_{j+1/2})}{\partial^{h,+}g_j}|\partial^{h,+}u_j(t)|^2].$$

Energy propagation of discrete waves can be described through the **principal symbol** :

$$\wp(x,t,\xi,\tau) := -g'(x)\rho(g(x))\tau^2 + 4\sin^2\left(\frac{\xi}{2}\right)\frac{\sigma(g(x))}{g'(x)} \tag{8}$$

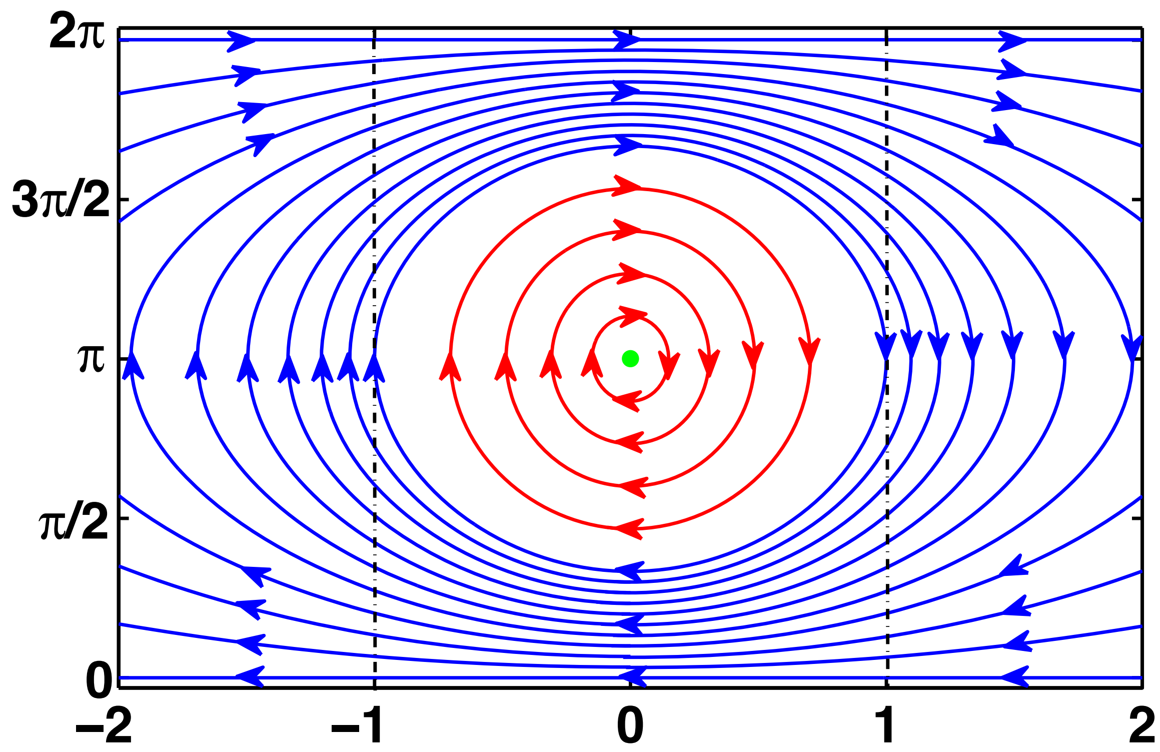
and the associated **null bi-characteristics**, solutions of the Hamiltonian system:

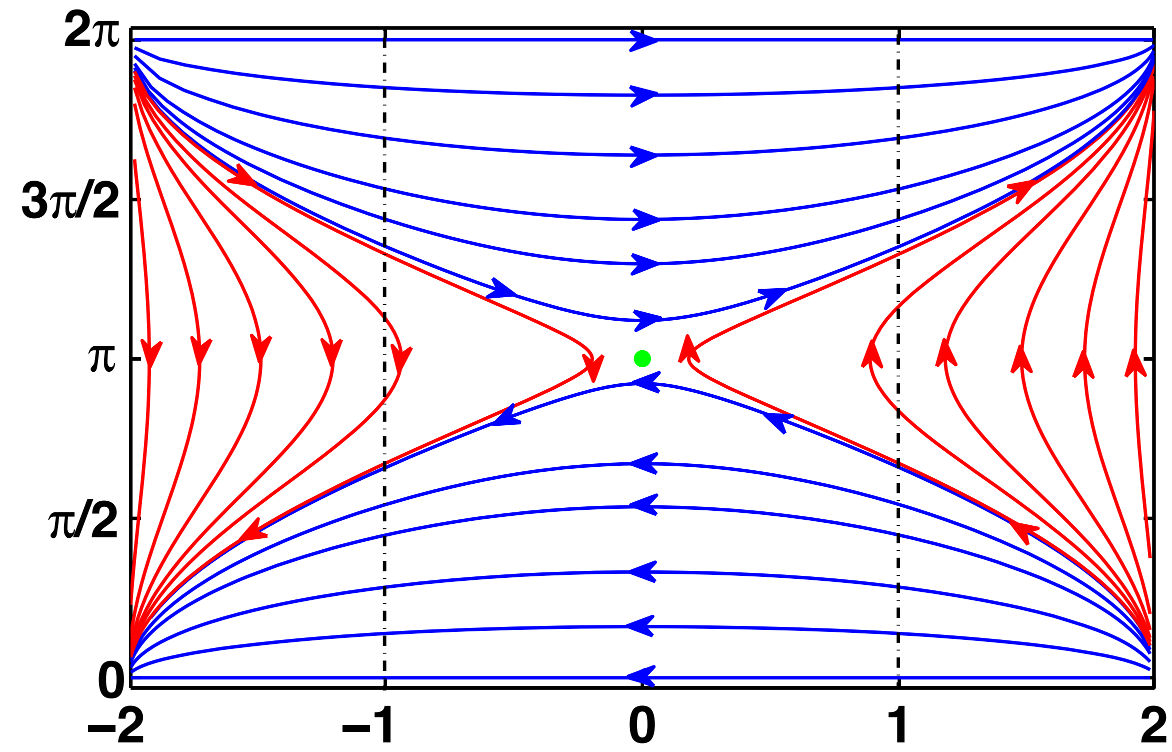
$$\begin{cases} \dot{X}(s) = \partial_\xi \wp = 2\sin(\Xi(s))\frac{\sigma(g(X(s)))}{g'(X(s))}, \\ \dot{t}(s) = \partial_\tau \wp = -g'(X(s))\rho(g(X(s)))\tau, \\ \dot{\Xi}(s) = -\partial_x \wp = (g'(\cdot)\rho(g(\cdot)))'(X(s)) - 4\sin^2\left(\frac{\Xi(s)}{2}\right)\left(\frac{\sigma(g(\cdot))}{g'(\cdot)}\right)'(X(s)), \\ \dot{\tau}(s) = -\partial_t \wp = 0, \end{cases} \tag{9}$$

$$(X)'(t) = \mp c_g(X(t))\cos\left(\frac{\Xi(t)}{2}\right),\ (\Xi)'(t) = \pm c_g'(X(t))2\sin\left(\frac{\Xi(t)}{2}\right) \tag{10}$$

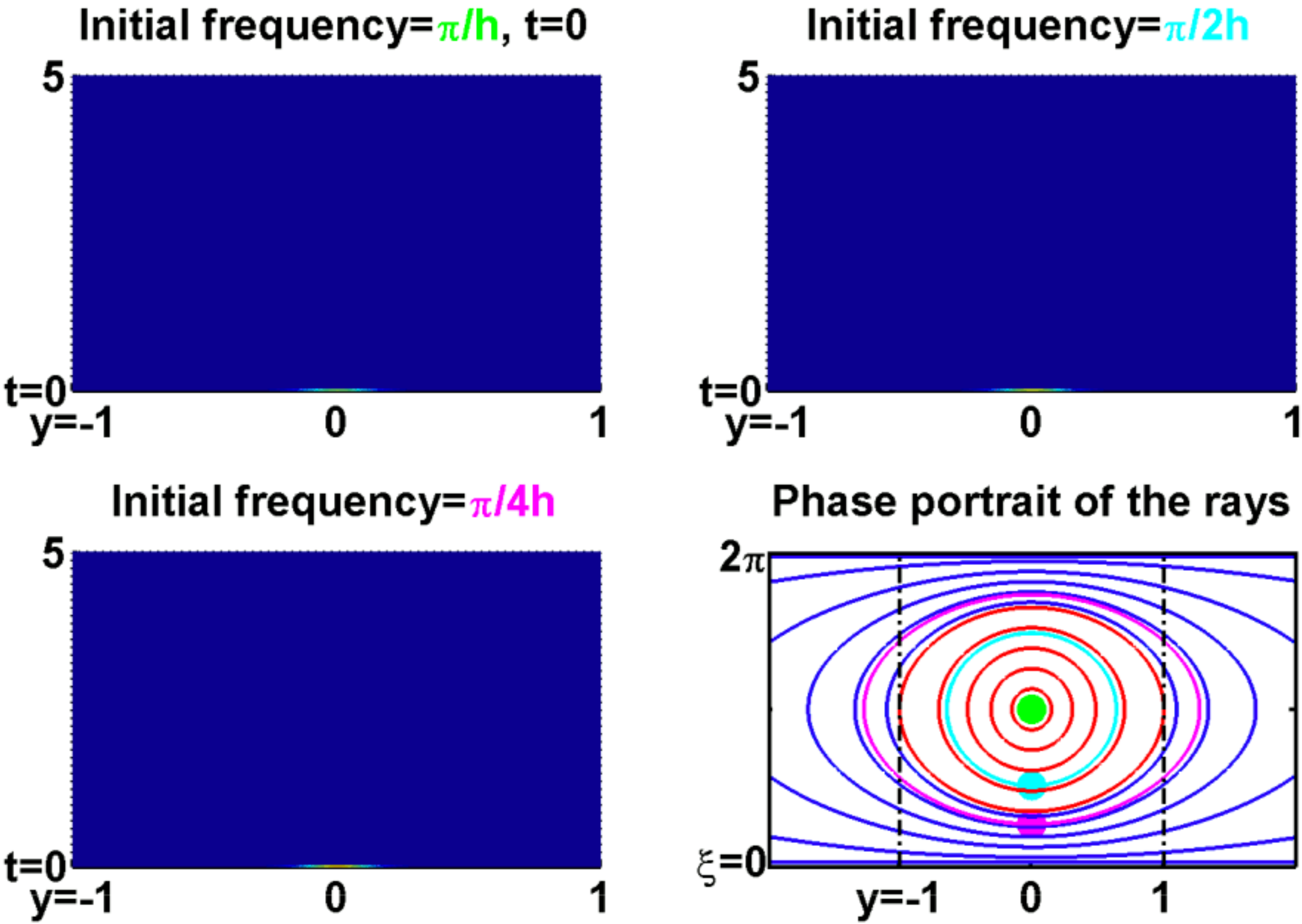
with

$$c_g(x) := \sqrt{\sigma(g(x))/\rho(g(x))}/g'(x).$$



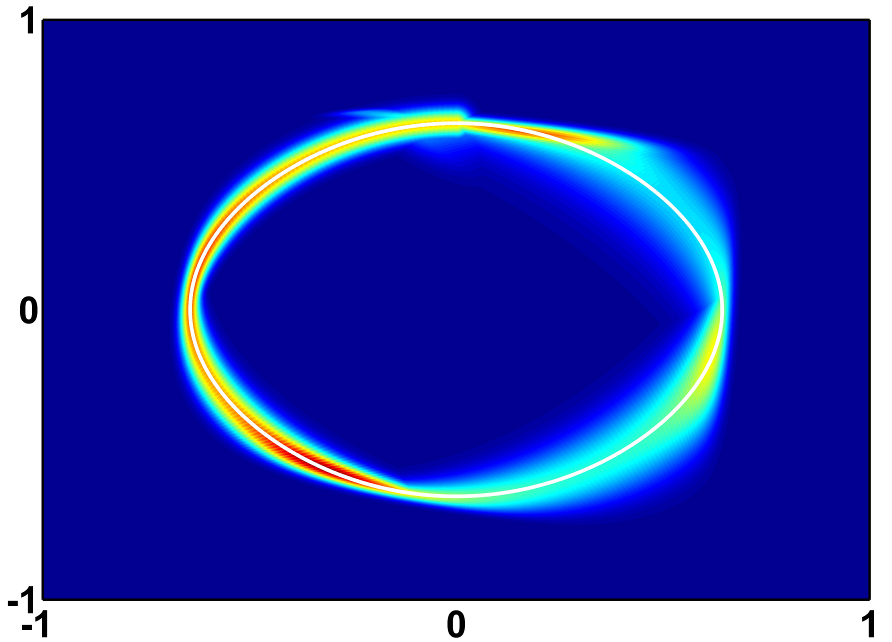


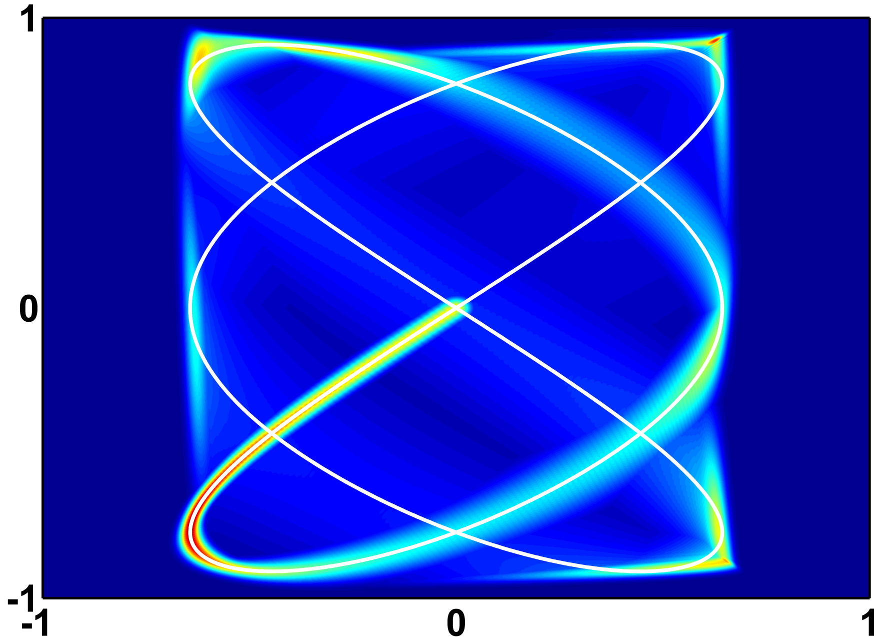
The phase portrait for the grid transformations $g(x) = \tan(\pi x/4)$ and $g(x) = 2\sin(\pi x/6)$.


Initial frequency=π/h, t=0
5
t=0
y=-1
0
1
Initial frequency=π/2h
5
t=0
y=-1
0
1
Initial frequency=π/4h
5
t=0
y=-1
0
1
Phase portrait of the rays
2π
ξ=0
y=-1
0
1

Multi-d “rodeo” spurious waves

(1) U. Biccari, A. Marica & E. Z., Propagation of one and two-dimensional discrete waves under finite difference approximation, FoCM (2020), 20:1401-1438.
(2) U. Biccari & E. Z., Gaussian Beam ansatz for finite difference wave equations, FoCM, 2023, 1-54.





Presentation Outline

- Numerical analysis is a mature field.
- Plenty remains to be done to assure that numerical schemes are qualitatively accurate.
- The effort needed is significant: Reproduce at the discrete level all the existing analysis techniques.
- Warning: The discrete world is full of ghosts!
- Ghosts proliferate as you move away from numerical analysis.