Uniform estimates for a fully discrete scheme integrating the linear heat equation on a bounded interval with pure Neumann boundary conditions

Guillaume Dujardin, Pauline Lafitte

Uniform estimates for a fully discrete scheme integrating the linear heat equation on a bounded interval with pure Neumann boundary conditions

Guillaume Dujardin[1] and Pauline Lafitte[2]

[1]Univ. Lille, INRIA, CNRS, UMR 8524 - Laboratoire Paul Painlevé F-59000
guillaume.dujardin@inria.fr
[2]Fédération de Mathématiques FR3487, Laboratoire MICS, CentraleSupélec, 3 rue Joliot-Curie, 91190 Gif-sur-Yvette pauline.lafitte@centralesupelec.fr

February 26, 2024

Abstract

This manuscript deals with the analysis of numerical methods for the full discretization (in time and space) of the linear heat equation with Neumann boundary conditions, and it provides the reader with error estimates that are uniform in time. First, we consider the homogeneous equation with homogeneous Neumann boundary conditions over a finite interval. Using finite differences in space and the Euler method in time, we prove that our method is of order 1 in space, uniformly in time, under a classical CFL condition, and despite its lack of consistency at the boundaries. Second, we consider the nonhomogeneous equation with nonhomogeneous Neumann boundary conditions over a finite interval. Using a tailored similar scheme, we prove that our method is also of order 1 in space, uniformly in time, under a classical CFL condition. We indicate how this numerical method allows for a new way to compute steady states of such equations when they exist. We conclude by several numerical experiments to illustrate the sharpness and relevance of our theoretical results, as well as to examine situations that do not meet the hypotheses of our theoretical results, and to illustrate how our results extend to higher dimensions.

1 Introduction

This article deals with the numerical integration of the classical linear heat equation in dimension 1, over a finite interval with homogeneous Neumann boundary conditions, and is concerned with proving *uniform in time* order estimates. The use of homogeneous Neumann boundary conditions is usual, for example when one wants to truncate an infinite domain (say, $\mathbb{R}$) to a finite one (say, a bounded interval) and one wants to “allow the solution to get out” (see for example [9] for the transport equation and references therein). Precisely, this article deals with methods using finite differences in space. This approach of using finite differences in space for solving linear PDEs in long time on bounded domains is simple and hence popular (see [11] for the wave equation and [9] for the transport equation), even though the treatment of the boundary condition raises delicate questions and sometimes unexpected behaviors. For example, it is well-known that, for the heat equation, the most naive classical finite-difference approach in space lacks consistency at the boundary of the domain [12]. Note however that, in this case, the associated matrix is symmetric and hence allows to carry out spectral analysis. A possible approach to circumvent this inconsistency issue is to use a modified finite-difference matrix on the first or last lines (as in [12] pages 21-23). This solution produces an order 1 scheme in space for the stationary problem that involves a nonsymmetric matrix. Another approach to circumvent this lack of consistency, developed in [19] (see page 15), consists in introducing a ghost point. Once again, this last approach yields a nonsymmetric finite-difference matrix.

In this paper, we consider the initial approach, with a symmetric matrix, and the associated time-dependent heat equation. For this discretization in space, and the explicit Euler method in time, despite this lack of consistency of the scheme at the boundary that we quantify, we manage to prove that the scheme has order $\mathcal{O}(\delta x)$ *uniformly in time*, under the classical CFL condition. To this end, we perform a thorough spectral analysis allowed by the symmetry of the finite-difference matrix. This is the main result of this paper and details are provided in Theorem 2.12. The proof relies on the one hand on a precise estimate of a consistency operator, and on the other hand on a precise control of the evolution of the numerical error. We use a discrete Gronwall lemma and both discrete and continuous coercivity estimates, that establish some uniform-in-time stability. This strategy is similar to the one used to prove the standard Lax theorem, which states that consistency and stability imply convergence. The originality of this paper is that we manage to carry out the analysis for all times, and we obtain error estimates that are *uniform in time*. To the best of our knowledge, this is the first ever uniform-in-time error analysis result for the heat equation with Neumann boundary conditions.

We also extend this analysis to address the *nonhomogeneous* linear heat equation (with a given source term and given fluxes at the boundary of the line segment). This provides a way to compute steady states of the heat equation with pure Neumann boundary conditions. The numerical approximation of the continuous operator in this context and its analysis arise *e.g.* in control problems (see for example [5], [4] and references therein). For the heat equation, the existence of such steady states is submitted to a compatibility condition between the heat fluxes at the boundary of the domain and the source term. When this compatibility condition is fulfilled, the direct computation of a steady state is known to be an ill-posed problem, because the continuous operator is nonnegative and self-adjoint with a nontrivial kernel. In this context, our method for the time integration of the time dependent heat equation can be seen as an iterative method to solve the corresponding discrete noninvertible linear problem. Our main result in this direction is Theorem 4.1. Solving this kind of problems numerically has a long history and may involve algebraic as well as variational formulations of the problem [1]. Other approaches use a Monte–Carlo formulation *via* a stochastic representation of the solution [13].

Most results about the convergence of numerical schemes for parabolic problems deal with finite time horizons (see for example a finite element methods for nonlinear heat equations with Dirichlet boundary conditions in [3], a Schwarz waveform relaxation method for the linear heat equation in [7], or numerical methods for fractional heat equations with non smooth data in [10]). Some results about numerical schemes for parabolic problems deal with the asymptotic behaviour of the schemes in time, but the question of the order of the scheme for all times is usually not addressed (see for example the long-time analysis of numerical methods for linear advection diffusion equations, using finite volume discretization for Dirichlet and Neumann boundary conditions in [2]). When the convergence of the scheme is addressed uniformly in time, it usually often for problems with Dirichlet boundary conditions and for weak (or weak-star) topologies. For example, in [18], the author considers two nonlinear Galerkin methods for the nonlinear Navier–Stokes system in a bounded domain Ω of $\mathbb{R}^2$ with homogeneous Dirichlet boundary conditions, and obtains convergence in $L^2(0, T, H)$ (where H is an appropriate subspace of $L^2(\Omega)^2$) for any *finite* time horizon $T > 0$ and weak-star convergence in $L^\infty(\mathbb{R}^+, H)$. In contrast, our aim is to obtain *uniform in time* strong estimates, to prove convergence and *uniform order* of our scheme, and to describe how one can handle Neumann boundary conditions using a discretization with finite differences that lacks consistency at the boundary of the domain. Some authors addressed a similar question of obtaining *uniform* convergence in strong topologies as well as *uniform* order estimates. For example, the present case of the nonlinear heat equation on a bounded interval with Dirichlet boundary condition using the forward Euler method in time and finite differences in space has been analyzed in [16]. In some sense, the sections 2 and 3 of this paper are the analogue of the case $f \equiv 0$ in [16] to the case of Neumann boundary conditions. Note that, anyway, frameworks for the longtime nonlinear analysis of schemes for parabolic equations exist (see for example [20]). However, they do not allow for the analysis of the scheme introduced in this paper (see Section 6.3), because of the way the Neumann conditions are discretized (lack of consistency with the Laplace operator at the boundary). Other authors proved convergence results for parabolic problems in the context of a data assimilation algorithm [14] [8] for the Navier–Stokes equation in dimension 2. In contrast, in this paper, we focus on the classical linear heat equation with Neumann boundary conditions on a bounded interval, we consider a fully discrete scheme based on finite differences, and we prove order estimates that are *uniform in time* for homogeneous as well as nonhomogeneous problems.

The main reason we address this classical problem is that it appears to be a very simplified version of the problem of the time integration of the linear Fokker–Planck equation (see for example [6]). In this setting, when the Fokker–Planck equation is homogeneous-in-space, the operator is symmetric nonnegative (with a nontrivial kernel), and coercivity estimates are a crucial tool to prove exponential convergence towards equilibrium, in the continuous and discrete settings. Moreover, this symmetric operator also lacks consistency at the boundary of the velocity domain. For this reason, we wish to consider discrete linear schemes with symmetric matrices for the operators in the linear continuous equation that are self-adjoint. For the Fokker–Planck equation, even in the homogeneous-in-space case, the analysis of the uniform order in time is still an open problem. The linear heat equation in this paper therefore serves as a toy model for this problem. Even if the analysis is carried out in dimension 1, our results extend to higher dimensions.

The outline of this paper is as follows. Section 2 is devoted to the introduction of the problem and the statement of the main result (Theorem 2.12). Section 3 deals with the estimation of the errors in time, and the proof of the main Theorem. Section 4 presents an application of this result to the computation of the steady state for nonhomogeneous Neumann problems. Numerical experiments are provided in Section 5 to show the efficiency and optimality of the theoretical results of the previous sections, and to illustrate the validity of these results in higher dimensions. This article ends with an appendix containing technical lemmas and a discussion about the possible generalization of the method to the discretization of the linear homogeneous Fokker–Planck equation with Neumann boundary conditions.

2 Setting of the problem and main results

2.1 The continuous linear heat equation with homogeneous Neumann boundary conditions

We consider the solution $u = u(t,x)$ to the problem

$$\begin{cases} \partial_t u(t,x) &= Pu(t,x) \\ u(0,x) &= u^0(x) \end{cases}, \tag{1}$$

where $P = \partial_x^2$ is the Laplace operator over $(0,L)$ with homogeneous Neumann boundary conditions at $x = 0$ and $x = L$, for some $L > 0$. The function $u^0 \in L^2(0,L)$ is some given initial datum, possibly smoother. We decompose the unknown solution $u(t,\cdot)$ of (1) at time $t > 0$ onto the classical orthonormal basis of $L^2(0,L)$ associated to P for the scalar product

$$\langle u, v \rangle = \frac{1}{L} \int_0^L u(x)v(x)\mathrm{d}x, \tag{2}$$

in the form

$$u(t,x) = \sum_{p=0}^{+\infty} \alpha_p \mathrm{e}^{-p^2 \frac{\pi^2}{L^2} t} c_p(x), \tag{3}$$

for $t \geq 0$ and $x \in (0, L)$, with

$$c_p(x) = \sqrt{2} \cos\left(p \frac{\pi}{L} x\right), \ p \geq 1, \qquad \text{and} \qquad c_0(x) = 1, \tag{4}$$

and $\alpha_p = \langle u^0, c_p \rangle$ for $p \geq 0$.

We denote by $1_{[0,L]}$ the constant function equal to 1 over $[0, L]$. We denote the norm associated to the scalar product (2) by $\| \cdot \|_{L^2}$ and the mean value of any function $v \in L^2(0, L)$ by

$$\langle v \rangle = \langle v, 1_{[0,L]} \rangle = \frac{1}{L} \int_0^L v(x) \mathrm{d}x. \tag{5}$$

2.2 Decay properties of the solutions of (1)

A classical result for the solutions to the linear heat equation (1) is the following

Property 2.1. *Assume $u^0 \in L^2(0, L)$. The corresponding solution u to* (1) *satisfies*

$$\forall t \geq 0, \qquad \|u(t, \cdot) - \langle u^0 \rangle 1_{[0,L]} \|_{L^2} \leq \mathrm{e}^{-\frac{\pi^2}{L^2} t} \|u^0\|_{L^2}.$$

Proof. Using twice the fact that the functions $(c_p)_{p \geq 0}$ are an orthonormal Hilbert basis of $L^2(0, L)$, we have, with the notations introduced above, for all $t \geq 0$,

$$\left\|u(t, \cdot) - \langle u^0 \rangle 1_{[0,L]}\right\|_{L^2}^2 = \sum_{p=1}^{+\infty} \left(\alpha_p \, \mathrm{e}^{-p^2 \frac{\pi^2}{L^2} t}\right)^2 \leq \mathrm{e}^{-2\frac{\pi^2}{L^2} t} \sum_{p=1}^{+\infty} \alpha_p^2 \leq \mathrm{e}^{-2\frac{\pi^2}{L^2} t} \left\|u^0\right\|_{L^2}^2.$$

□

Assuming extra smoothness on the initial datum u^0, this decay property also holds for x-derivatives of the exact solution to the linear equation (1), as is stated, for example, in Property 2.6.

Property 2.2 (Domain of the operator P)**.** *The domain* $\mathrm{Dom}(P)$ *of the operator P is the set of functions $u \in L^2(0, L)$ such that $\partial_x^2 u \in L^2(0, L)$ and $\partial_x u(0) = \partial_x u(L) = 0$. Observe in particular that* $\mathrm{Dom}(P) \subset H^2(0, L)$.

Throughout the paper, we assume the following on the initial datum u^0:

$$u^0 \in \mathrm{Dom}(P), \qquad P u^0 \in \mathrm{Dom}(P), \qquad P^2 u^0 \in \mathrm{Dom}(P). \tag{6}$$

Remark 2.3. *The hypothesis* (6) *is for example fulfilled for example in either of the following cases:*

- *u^0 is $\mathcal{C}^\infty$ over $(0, L)$ with compact support,*
- *u^0 is the restriction to $[0, L]$ of an even $2L$-periodic function of class $\mathcal{C}^6$ over $\mathbb{R}$.*

Remark 2.4. *Assume that* $u^0 \in L^2(0,L)$*. Then, for all* $t > 0$*, the corresponding solution* $u(t) = \mathrm{e}^{-tP}u^0$ *of* (1) *at time* t *defined in* (3) *satisfies the hypothesis* (6)*. In particular, if* $u^0 \in L^2(0,L)$ *does not satisfy the hypothesis* (6)*, then it does instantaneously after* $t = 0$*. The numerical long-time analysis presented below can surely be adapted using this remark to suppress the hypothesis* (6) *and replace it with the simple hypothesis that* $u^0 \in L^2(0,L)$ *(see for example Remark 4.3). We shall not do this here for the sake of brevity. However, we illustrate this fact numerically in Section 5.*

Property 2.5. *If* u^0 *satisfies* (6)*, then the corresponding solution* u *of* (1) *defined in* (3) *is in* $\mathcal{C}^2([0;+\infty[, L^2(0,L))$*.*

Property 2.6. *Assume that* u^0 *satisfies* (6)*. Then, the corresponding solution* u *to* (1) *obtained by* (3) *satisfies*

$$\forall t \geq 0, \qquad \left\|P^2 u(t)\right\|_{L^2} \leq \mathrm{e}^{-\frac{\pi^2}{L^2}t} \left\|P^2 u^0\right\|_{L^2}, \tag{7}$$

and

$$\forall t \geq 0, \qquad \left\|\partial_x P^2 u(t)\right\|_{L^2} \leq \mathrm{e}^{-\frac{\pi^2}{L^2}t} \left\|\partial_x P^2 u^0\right\|_{L^2}. \tag{8}$$

Proof. Since $P^2u^0 \in \mathrm{Dom}(P)$, we have that $t \mapsto P^2u(t)$ is continuous from $[0;+\infty)$ to $L^2(0,L)$. Since it has zero mean value and solves the homogeneous linear heat equation with homogeneous Neumann boundary condition, with initial datum $P^2u^0 \in L^2(0,L)$ with zero mean value (since $Pu^0 \in \mathrm{Dom}(P)$), we infer that (7) holds, using Property 2.1.

Moreover, the function $t \mapsto \partial_x^5 u(t)$ is continuous over $[0,+\infty)$ with values in $L^2(0,L)$, and solves the linear homogeneous heat equation with homogeneous Dirichlet boundary condition, and initial datum $\partial_x^5 u^0 \in L^2(0,L)$ (which vanishes at 0 and L since $P^2u^0 \in \mathrm{Dom}(P)$). We infer that (8) holds. □

2.3 The discretized problem

We consider the discretization of the Laplace operator with homogeneous Neumann boundary conditions on $(0,L)$ that appears in (1). To do so, we set $J \geq 2$ and define $x_j = j\delta x$ for $0 \leq j \leq J-1$ and $\delta x = L/(J-1)$. We equip $\mathbb{R}^J$ with the inner product

$$\langle \mathsf{v}, \mathsf{w}\rangle_\delta = \frac{1}{J}\sum_{j=0}^{J-1} \mathsf{v}_j \mathsf{w}_j, \qquad (\mathsf{v},\mathsf{w}) \in \mathbb{R}^J \times \mathbb{R}^J, \tag{9}$$

which is a discrete analogue to (2), and denote by $\|\cdot\|_{\ell^2}$ the associated norm. And, similarly to (5), denoting $1 = (1,\dots,1)^\top$, we set

$$\langle \mathsf{v}\rangle_\delta = \langle \mathsf{v}, 1\rangle_\delta = \frac{1}{J}\sum_{j=0}^{J-1} \mathsf{v}_j, \tag{10}$$

for all $\mathsf{v} \in \mathbb{R}^J$. We can think of v_j as an approximation of some smooth function v on $[0,L]$ at point x_j. We introduce the classical matrix

$$\mathsf{P}_\delta = \frac{1}{\delta x^2}\begin{pmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & -2 & 1 \\ 0 & \dots & \dots & 0 & 1 & -1 \end{pmatrix}.$$

It is a real symmetric matrix and P_δ has J simple real eigenvalues which read

$$\lambda_\ell = -\frac{4}{\delta x^2}\sin^2\left(\frac{\ell\pi}{2J}\right), \qquad 0 \le \ell \le J-1. \tag{11}$$

The eigenspace for the eigenvalue $\lambda_0 = 0$ is generated by the vector 1. See Section 3.1 for more on the spectral analysis of P_δ.

For $\delta t > 0$ we consider the explicit fully discrete iterative scheme

$$\begin{cases} \mathsf{v}^{n+1} & = & (\mathsf{Id} + \delta t \mathsf{P}_\delta)\mathsf{v}^n \\ \mathsf{v}^0 & & \text{given in } \mathbb{R}^J \end{cases}, \tag{12}$$

which is a discrete analogue of (1). The classical CFL condition of this scheme (see Proposition 2.7) reads

$$\frac{\delta t}{\delta x^2} \le \frac{1}{2}. \tag{13}$$

Note that (12) is a fully discrete analogue of (1), where P is discretized by P_δ in space and the explicit Euler method is used in time. We can think of v^0 as a discrete approximation of u^0. Let us denote by $\|\cdot\|_{\ell^2}$ the norm associated to the scalar product defined in (9). We have the following classical stability result.

Proposition 2.7. *Assume $J \ge 2$ and $\delta t > 0$ are such that the CFL condition* (13) *is fulfilled. We have*

$$\rho((\mathsf{Id} + \delta t \mathsf{P}_\delta)) = 1 \qquad \textit{and} \qquad \forall \mathsf{v} \in \mathbb{R}^J, \quad \|(\mathrm{Id} + \delta \mathrm{t} \mathsf{P}_\delta)\mathsf{v}\|_{\ell^2} \le \|\mathsf{v}\|_{\ell^2}, \tag{14}$$

where ρ denotes the spectral radius.

Proof. The CFL condition (13) ensures that the eigenvalues of P_δ satisfy (see Equation 11)

$$\forall \ell \in \{0, \cdots, J-1\}, \qquad |1 + \delta t \lambda_\ell| \le 1.$$

This implies that $\rho(\mathsf{Id} + \delta t \mathsf{P}_\delta) = 1$ (recall that $\lambda_0 = 0$). The fact that, additionally, P_δ is symmetric and has eigenvectors forming an orthogonal basis of $\mathbb{R}^J$ yields (14). $\square$

We make repeated use of the existence of a spectral gap for the operator P_δ, which is described in the following proposition.

Proposition 2.8. *Assume $L > 0$ is fixed, $J \ge 2$ and $\delta t > 0$ are given. We set, using the eigenvalues $(\lambda_\ell)_{0\le\ell\le J-1}$ of P_δ defined in* (11),

$$\eta = \max_{1\le\ell\le J-1} |1 + \delta t \lambda_\ell|\,. \tag{15}$$

Assuming that the CFL condition (13) *is fulfilled, we have*

$$0 < \eta < 1. \tag{16}$$

Proof. Thanks to (11), we observe that $\lambda_0 = 0$ and for $\ell \in \{1, \cdots, J-1\}$, $0 < \ell\pi/(2J) < \pi/2$, so that $0 < \sin^2(\ell\pi/(2J)) < 1$ and hence, using (13), $-2 < -4\delta t/(\delta x^2) \times \sin^2(\ell\pi/(2J)) < 0$, and therefore

$$-1 < 1 + \delta t \lambda_\ell < 1.$$

This proves (16), since the $J-1$ values $(1 + \delta t \lambda_\ell)_{1\le\ell\le J-1}$ are all distinct (hence not all zero) as soon as $J \ge 3$ ($J = 2$ is easily treated separately). $\square$

Another useful and more precise estimate on the eigenvalues of $I + \delta t \mathsf{P}_\delta$ is provided in Proposition 3.2.

2.4 Description of the lack of consistency

Definition 2.9. *We introduce the "projection" operator $\Pi_{\delta x}$ acting on continuous functions w over the closed bounded interval $[0, L]$ by setting for all $j \in \{0, \cdots, J-1\}$, $(\Pi_{\delta x}(w))_j = w(x_j)$.*

The consistency of the operator P_δ with respect to P can be measured using the following operator.

Definition 2.10. *For all smooth function w over $[0, L]$, we set for all $J \geq 2$,*

$$\mathcal{L}_\delta w = (\Pi_{\delta x} P - \mathsf{P}_\delta \Pi_{\delta x})\, w = \begin{pmatrix} \partial_x^2 w(x_0) - \frac{w(x_1) - w(x_0)}{\delta x^2} \\ \partial_x^2 w(x_1) - \frac{w(x_0) - 2w(x_1) + w(x_2)}{\delta x^2} \\ \vdots \\ \partial_x^2 w(x_{J-2}) - \frac{w(x_{J-3}) - 2w(x_{J-2}) + w(x_{J-1})}{\delta x^2} \\ \partial_x^2 w(x_{J-1}) - \frac{w(x_{J-2}) - w(x_{J-1})}{\delta x^2} \end{pmatrix}.$$

For the analysis to come, we split the consistency defect above into two terms.

Proposition 2.11. *For all smooth function w over $[0, L]$ such that $\partial_x w(x_0) = \partial_x w(x_{J-1}) = 0$, we have for all $J \geq 2$,*

$$\mathcal{L}_\delta w := (\Pi_{\delta x} P - \mathsf{P}_\delta \Pi_{\delta x})\, w = \mathcal{L}_\delta^1 w + \delta x^2 \mathcal{L}_\delta^2 w, \tag{17}$$

with

$$\mathcal{L}_\delta^1 w = \begin{pmatrix} \frac{1}{2}\partial_x^2 w(x_0) + \frac{\delta x}{2} \int_0^1 (1-\sigma)^2 \partial_x^3 w(\sigma \delta x) \mathrm{d}\sigma \\ 0 \\ \vdots \\ 0 \\ \frac{1}{2}\partial_x^2 w(x_{J-1}) + \frac{\delta x}{2} \int_0^1 (1-\sigma)^2 \partial_x^3 w(x_{J-2} + \sigma \delta x) \mathrm{d}\sigma \end{pmatrix}, \tag{18}$$

and

$$\mathcal{L}_\delta^2 w = \frac{1}{6} \begin{pmatrix} 0 \\ \int_0^1 (1-\sigma)^3 \partial_x^4 w(x_1 + \sigma \delta x) \mathrm{d}\sigma + \int_0^1 (\sigma - 1)^3 \partial_x^4 w(x_1 - \sigma \delta x) \mathrm{d}\sigma \\ \vdots \\ \int_0^1 (1-\sigma)^3 \partial_x^4 w(x_{J-2} + \sigma \delta x) \mathrm{d}\sigma + \int_0^1 (\sigma - 1)^3 \partial_x^4 w(x_{J-2} - \sigma \delta x) \mathrm{d}\sigma \\ 0 \end{pmatrix}. \tag{19}$$

An interpretation of the splitting (17) of the consistency error of the operator P_δ with respect to P is the following: The operator P_δ is consistent with the operator P to order δx^2 at interior points $(x_j)_{1 \leq j \leq J-2}$, yet it is *not* consistent at boundary points x_0 and x_{J-1}. This is the main difficulty in the error analysis of the scheme (12). We shall explain how to deal with this lack of consistency in Section 3.2.

2.5 Main results

The main result of this paper is the following error estimate, that is valid *uniformly in time* for the approximation of the solutions of the linear heat equation (1) by the linear scheme (12). In particular, this result encapsulates the lack of consistency that appears through the operator $\mathcal{L}_{\delta x}^1 + \delta x^2 \mathcal{L}_{\delta x}^2$ defined in (17), and shows how the error behaves over short and long times. The proof of the result is carried out in Section 3, and numerical results illustrating and supporting it are provided in Section 5.

Theorem 2.12. *Assume $L > 0$ is fixed. There exists $C > 0$ such that for all $u^0 \in H^6(0, L)$ satisfying* (6)*, for all $J \geq 2$, $\delta t \in (0, 1)$ such that the CFL condition* (13) *holds, and all $n \geq 1$,*

$$\|\Pi_{\delta x} u(n\delta t) - \mathsf{v}^n\|_{\ell^2} \leq \left\|\Pi_{\delta x} u^0 - \mathsf{v}^0\right\|_{\ell^2} + C\left(\delta x \sum_{p=1}^{+\infty} |\alpha_p| p^4 + \delta t \left\|P^2 u^0\right\|_{H^1}\right). \tag{20}$$

This theorem implies straightforwardly the following corollary.

Corollary 2.13. *Assume $L > 0$ is fixed. There exists $C > 0$ such that for all $u^0 \in H^6(0, L)$ satisfying* (6)*, for all $J \geq 2$, $\delta t \in (0, 1)$ such that the CFL condition* (13) *holds, the numerical solution $(\mathsf{v}^n)_{n\geq 0}$ of* (12) *starting from $\mathsf{v}^0 = \Pi_{\delta x} u^0$ satisfies for all $n \geq 1$,*

$$\|\Pi_{\delta x} u(n\delta t) - \mathsf{v}^n\|_{\ell^2} \leq C(\delta x + \delta t)\|u^0\|_{H^5}.$$

Proof. On the one hand, since u^0 satisfies (6), it is obvious that

$$\left\|P^2 u^0\right\|_{H^1} \leq \|u^0\|_{H^5}.$$

On the other hand, we have

$$\sum_{p=1}^{+\infty} |\alpha_p| p^4 \leq \left(\sum_{p=1}^{+\infty} |\alpha_p|^2 p^{10}\right)^{\frac{1}{2}} \times \left(\sum_{p=1}^{+\infty} p^{-2}\right)^{\frac{1}{2}} \leq C\left(\sum_{p=1}^{+\infty} |\alpha_p|^2 \left(\frac{p\pi}{L}\right)^{10}\right)^{\frac{1}{2}} \leq C\|\partial_x^5 u^0\|_{L^2} \leq C\|u^0\|_{H^5},$$

where we used (6) for the computation of the coefficients of $\partial_x^5 u^0$ in the sine basis of $L^2(0, L)$. This concludes the proof of the corollary. □

These two results say, in particular, that the scheme (12) applied to the discretized version of (1) using the finite-difference matrix P_δ, under the CFL condition (13), has *uniform-in-time* order $\mathcal{O}(\delta t^{1/2})$, if we have in mind that, in the specific CFL regime $\delta t = \delta x^2/2$, $\delta x = \mathcal{O}(\delta t^{\frac{1}{2}})$.

3 Error analysis

This section is devoted to the proof of Theorem 2.12. Section 3.1 extends the spectral analysis of the operator P_δ. Sections 3.2 and 3.3 present an analysis of the second and third terms in the decomposition (21) of the error below. Finally, Section 3.4 sums up the previous result, and allows for the proof of Theorem 2.12.

Using the operator $\Pi_{\delta x}$ defined above (see Definition 2.9), we define for all $n \in \mathbb{N}$ the convergence error at time step number n by

$$\mathsf{e}^n = \Pi_{\delta x} u(n\delta t) - \mathsf{v}^n,$$

where u solves (1) and $(\mathsf{v}^n)_{n\geq 0}$ is given by (12). Since u^0 satisfies (6), Property 2.5 ensures that $t \mapsto u(t)$ is $\mathcal{C}^2$ from $[0, +\infty)$ to $L^2(0, L)$. Hence, using (1) in a Taylor expansion in integral remainder form of $t \mapsto u(t)$ at $t = n\delta t$, we obtain

$$\begin{aligned}
\mathsf{e}^{n+1} &= \Pi_{\delta x}\left(u(n\delta t) + \delta t P u(n\delta t) + \int_{n\delta t}^{(n+1)\delta t} ((n+1)\delta t - s) P^2 u(s)\mathrm{d}s\right) - (\mathsf{Id} + \delta t \mathsf{P}_\delta)\,\mathsf{v}^n \\
&= \Pi_{\delta x} u(n\delta t) + \delta t \Pi_{\delta x} P u(n\delta t) - (\mathsf{Id} + \delta t \mathsf{P}_\delta)\,\mathsf{v}^n + \Pi_{\delta x}\int_{n\delta t}^{(n+1)\delta t} ((n+1)\delta t - s) P^2 u(s)\mathrm{d}s \\
&= (\mathsf{Id} + \delta t \mathsf{P}_\delta)\,\mathsf{e}^n + \underbrace{\delta t\,(\Pi_{\delta x} P - \mathsf{P}_\delta \Pi_{\delta x})\,u(n\delta t)}_{:=\varepsilon_1^n} + \underbrace{\int_{n\delta t}^{(n+1)\delta t} ((n+1)\delta t - s)\Pi_{\delta x} P^2 u(s)\mathrm{d}s}_{:=\varepsilon_2^n}.
\end{aligned}$$

This yields the expression of the error as

$$\mathsf{e}^n = (\mathsf{Id} + \delta t \mathsf{P}_\delta)^n\,\mathsf{e}^0 + \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k}\,\varepsilon_1^k + \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k}\,\varepsilon_2^k. \tag{21}$$

Observe that the terms in ε_1 and ε_2 only depend on the exact solution u. Note also that ε_1 contains a factor δt and ε_2 scales as δt^2.

The goal of this section is to establish uniform in time estimates on the three terms in the right-hand side of (21), in order to prove Theorem 2.12. The term with ε_1 is studied in Section 3.2 and that with ε_2 is studied in Section 3.3. We start in Section 3.1 with some additional spectral properties of the matrix P_δ.

3.1 Spectral analysis of P_δ

Spectral decomposition of P_δ

Lemma 3.1. *For all $J \geq 2$, the symmetric matrix P_δ is nonpositive. Its eigenvalues are simple and given by* (11)*. Moreover, the corresponding eigenvectors $\mathsf{W}_0 = 1$ and, for $\ell \in \{1, \cdot, J-1\}$,*

$$(\mathsf{W}_\ell)_j = \sqrt{2}\cos(\ell(j+1/2)\pi/J), \qquad j \in \{0, \dots, J-1\},$$

form an orthonormal basis of $\mathbb{R}^J$ for the inner product defined in (9)*.*

The proof, which is very classical, is not included in this paper.

Decomposition of the boundary terms using the spectral decomposition of P_δ

Let us denote by $(e_j)_{0\leq j\leq J-1}$ the canonical basis of $\mathbb{R}^J$. Using the spectral decomposition of P_δ given by Lemma 3.1, we have

$$e_0 = \sum_{\ell=0}^{J-1} \langle e_0, \mathsf{W}_\ell\rangle_\delta \mathsf{W}_\ell = \frac{\sqrt{2}}{J}\sum_{\ell=1}^{J-1} \cos\left(\frac{\ell\pi}{2J}\right)\mathsf{W}_\ell + \frac{1}{J}\mathsf{W}_0, \tag{22}$$

and

$$e_{J-1} = \sum_{\ell=0}^{J-1} \langle e_{J-1}, \mathsf{W}_\ell\rangle_\delta \mathsf{W}_\ell = \frac{\sqrt{2}}{J}\sum_{\ell=1}^{J-1} (-1)^\ell \cos\left(\frac{\ell\pi}{2J}\right)\mathsf{W}_\ell + \frac{1}{J}\mathsf{W}_0. \tag{23}$$

Estimates of the powers of $(\mathsf{Id} + \delta t \mathsf{P}_\delta)$

Proposition 3.2. *For all* $L > 0$, $\delta t > 0$ *and* $J \geq 2$ *such that* (13) *holds, one has for all* $\ell \in \{0, \cdots, J-1\}$,

$$|1 + \delta t \lambda_\ell| \leq \mathrm{e}^{-\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{J}\right)^2}. \tag{24}$$

Proof. Observe that, for $\ell \in \{0, \cdots, J-1\}$, we have

$$1 + \delta t \lambda_\ell = 1 - 4\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2.$$

In particular,

$$(1 + \delta t \lambda_\ell)^2 = 1 - 8\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2 + 16\frac{\delta t^2}{\delta x^4} \sin\left(\frac{\ell\pi}{2J}\right)^4.$$

Using that for all $u \in \mathbb{R}$, $1 - u \leq \mathrm{e}^{-u}$, we obtain

$$\begin{aligned}
(1 + \delta t \lambda_\ell)^2 &\leq \exp\left(-8\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2 + 16\frac{\delta t^2}{\delta x^4} \sin\left(\frac{\ell\pi}{2J}\right)^4\right) \\
&\leq \exp\left(-8\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2 \left(1 - 2\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2\right)\right) \\
&\leq \exp\left(-8\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2 \left[\left(1 - 2\frac{\delta t}{\delta x^2}\right) \times 1 + 2\frac{\delta t}{\delta x^2}\left(1 - \sin\left(\frac{\ell\pi}{2J}\right)^2\right)\right]\right).
\end{aligned}$$

In particular, taking square roots,

$$|1 + \delta t \lambda_\ell| \leq \exp\left(-4\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2 \left[\left(1 - 2\frac{\delta t}{\delta x^2}\right) \times 1 + 2\frac{\delta t}{\delta x^2}\left(1 - \sin\left(\frac{\ell\pi}{2J}\right)^2\right)\right]\right).$$

Using (13), we have

$$2\frac{\delta t}{\delta x^2} \geq 0, \qquad 1 - 2\frac{\delta t}{\delta x^2} \geq 0,$$

and the sum of these two real numbers is 1. By convexity over $\mathbb{R}$ of the exponential function, we infer

$$\begin{aligned}
&|1 + \delta t \lambda_\ell| \\
&\leq \left(1 - 2\frac{\delta t}{\delta x^2}\right) \exp\left(-4\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2 \times 1\right) + 2\frac{\delta t}{\delta x^2} \exp\left(-4\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2 \times \left(1 - \sin\left(\frac{\ell\pi}{2J}\right)^2\right)\right) \\
&\leq \left(1 - 2\frac{\delta t}{\delta x^2}\right) \exp\left(-4\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2 \times 1\right) + 2\frac{\delta t}{\delta x^2} \exp\left(-4\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2 \times \cos\left(\frac{\ell\pi}{2J}\right)^2\right) \\
&\leq \left(1 - 2\frac{\delta t}{\delta x^2}\right) \exp\left(-4\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2 \times \cos\left(\frac{\ell\pi}{2J}\right)^2\right) + 2\frac{\delta t}{\delta x^2} \exp\left(-4\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2 \times \cos\left(\frac{\ell\pi}{2J}\right)^2\right) \\
&\leq \exp\left(-4\frac{\delta t}{\delta x^2} \sin\left(\frac{\ell\pi}{2J}\right)^2 \times \cos\left(\frac{\ell\pi}{2J}\right)^2\right).
\end{aligned}$$

As a conclusion, this implies (24). $\square$

We provide the following estimate of the sum of the powers of $(\mathsf{Id} + \delta t \mathsf{P}_\delta)$.

Proposition 3.3. *Assume $L > 0$ is fixed, $J \geq 2$ and $\delta t > 0$ are given. Let η be defined by* (15)*. Assuming the CFL condition* (13)*, we have*

$$\forall n \geq 1, \qquad \delta t \sum_{k=0}^{n-1} \eta^k \leq 2L^2.$$

Proof. On the one hand, using the CFL condition (13), Proposition 3.2 and the definition (15) of η ensure that $\eta \leq \mathrm{e}^{-\frac{\delta t}{\delta x^2}\sin\left(\frac{\pi}{J}\right)^2} < 1$. This ensures that, for all $n \geq 1$,

$$\delta t \sum_{k=0}^{n-1} \eta^k \leq \delta t \sum_{k=0}^{n-1} \mathrm{e}^{-k\frac{\delta t}{\delta x}\sin\left(\frac{\pi}{J}\right)^2} \leq \delta t \frac{1 - \mathrm{e}^{-n\frac{\delta t}{\delta x^2}\sin\left(\frac{\pi}{J}\right)^2}}{1 - \mathrm{e}^{-\frac{\delta t}{\delta x^2}\sin\left(\frac{\pi}{J}\right)^2}} \leq \delta t \frac{1}{1 - \mathrm{e}^{-\frac{\delta t}{\delta x^2}\sin\left(\frac{\pi}{J}\right)^2}}.$$

Moreover, using the CFL condition (13), we have $\delta t \sin(\pi/J)^2/\delta x^2 \leq 1/2$, and hence this number $z > 0$ satisfies $1/(1 - \mathrm{e}^{-z}) \leq 2/z$. This implies that, for all $n \geq 1$,

$$\delta t \sum_{k=0}^{n-1} \eta^k \leq \delta t \frac{2}{\frac{\delta t}{\delta x^2}\sin\left(\frac{\pi}{J}\right)^2} \leq 2 \frac{\delta x^2}{\sin\left(\frac{\pi}{J}\right)^2} \leq 2 \frac{\pi^2}{4} \frac{\delta x^2}{\left(\frac{\pi}{J}\right)^2}.$$

Since $\delta x = L/(J-1)$, we have

$$\delta t \sum_{k=0}^{n-1} \eta^k \leq \frac{1}{2}(J\delta x)^2 \leq \frac{1}{2} L^2 \left(\frac{J}{J-1}\right)^2 \leq 2L^2,$$

since $J \geq 2$.

$\square$

Another useful estimate of the powers of $(\mathsf{Id} + \delta t \mathsf{P}_\delta)$ is the following, the proof of which can be found in Section 6.1.

Proposition 3.4. *For all $L > 0$, there exists $C > 0$ such that for all $J \geq 2$, for all $\delta t > 0$ such that* (13) *holds, we have*

$$\sum_{\ell=1}^{J-1} \left| \delta t \sum_{k=0}^{n-1} (1 + \delta t \lambda_\ell)^k \right|^2 \leq C,$$

where $(\lambda_\ell)_{0 \leq \ell \leq J-1}$ are the eigenvalues of P_δ (see (11)*).*

3.2 Analysis of the term in ε_1

We insert the splitting (17) into the error term with ε_1 in (21). We obtain

$$\begin{aligned}
\sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} \varepsilon_1^n &= \delta t \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} \mathcal{L}_\delta u(k\delta t, \cdot) \\
&= \delta t \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} \mathcal{L}_\delta^1 u(k\delta t, \cdot) + \delta t \delta x^2 \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} \mathcal{L}_\delta^2 u(k\delta t, \cdot).
\end{aligned}$$

In this section, we deal with the first and second term in this decomposition, and conclude.

Analysis of the term with $\mathcal{L}^1_\delta$

Proposition 3.5. *Assume $L>0$ is fixed. There exists $C>0$ such that for all $u^0 \in H^6(0,L)$ satisfying (6), for all $\delta t \in (0,1)$ and $J \geq 2$ such that the CFL condition (13) holds, and all $n \geq 1$,*

$$\left\| \delta t \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} \mathcal{L}^1_\delta u(k\delta t, \cdot) \right\|_{\ell^2} \leq C \delta x \sum_{p=1}^{+\infty} |\alpha_p| p^3. \tag{25}$$

Proof. Using the formula (3) for the exact solution u to (1), we obtain the following ℓ^2 error inequality,

$$\left\| \delta t \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} \mathcal{L}^1_\delta u(k\delta t, \cdot) \right\|_{\ell^2} \leq \sum_{p=1}^{+\infty} |\alpha_p| \left\| \delta t \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} (\mathcal{L}^1_\delta c_p) \mathrm{e}^{-p^2 \frac{\pi^2}{L^2} k \delta t} \right\|_{\ell^2}. \tag{26}$$

Observe the disappearance of the term in $p=0$ in the sum above, since $\mathcal{L}^1_{\delta x} c_0 = 0$ according to (18). Let us fix $p \geq 1$. Since the only possibly nonzero coefficients of $\mathcal{L}^1_\delta c_p$ are its first and last ones (see (18)), we can use the decompositions (22) and (23) in the orthonormal basis of $\mathbb{R}^J$ consisting in $\mathsf{W}_0, \cdots, \mathsf{W}_{J-1}$ (see Lemma 3.1):

$$\begin{aligned}
&\delta t \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} \left(\mathcal{L}^1_\delta c_p\right) \mathrm{e}^{-p^2 \frac{\pi^2}{L^2} k \delta t} \\
&= \frac{\sqrt{2}}{J} \delta t \sum_{\ell=1}^{J-1} \cos\left(\frac{\pi \ell}{2J}\right) \left(\left[\mathcal{L}^1_\delta c_p\right]_0 + (-1)^\ell \left[\mathcal{L}^1_\delta c_p\right]_{J-1} \right) \left(\sum_{k=0}^{n-1} (1 + \delta t \lambda_\ell)^{n-1-k} \mathrm{e}^{-p^2 \frac{\pi^2}{L^2} k \delta t} \right) \mathsf{W}_\ell \\
&\quad + \frac{\delta t}{J} \left(\left[\mathcal{L}^1_\delta c_p\right]_0 + \left[\mathcal{L}^1_\delta c_p\right]_{J-1} \right) \left(\sum_{k=0}^{n-1} \mathrm{e}^{-p^2 \frac{\pi^2}{L^2} k \delta t} \right) \mathsf{W}_0.
\end{aligned} \tag{27}$$

Using the orthonormality of the basis $(\mathsf{W}_\ell)_{0 \leq \ell \leq J-1}$, we obtain

$$\begin{aligned}
&\left\| \delta t \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} \mathcal{L}^1_\delta c_p \, \mathrm{e}^{-p^2 \frac{\pi^2}{L^2} k \delta t} \right\|_{\ell^2}^2 \\
&= \frac{2}{J^2} \delta t^2 \sum_{\ell=1}^{J-1} \cos^2\left(\frac{\pi \ell}{2J}\right) \left(\left[\mathcal{L}^1_\delta c_p\right]_0 + (-1)^\ell \left[\mathcal{L}^1_\delta c_p\right]_{J-1} \right)^2 \left(\sum_{k=0}^{n-1} (1 + \delta t \lambda_\ell)^{n-1-k} \mathrm{e}^{-p^2 \frac{\pi^2}{L^2} k \delta t} \right)^2 \\
&\quad + \frac{1}{J^2} \delta t^2 \left(\left[\mathcal{L}^1_\delta c_p\right]_0 + \left[\mathcal{L}^1_\delta c_p\right]_{J-1} \right)^2 \left(\sum_{k=0}^{n-1} \mathrm{e}^{-p^2 \frac{\pi^2}{L^2} k \delta t} \right)^2 \\
&\leq 4 \frac{\delta t^2}{J^2} \left(\left[\mathcal{L}^1_\delta c_p\right]_0^2 + \left[\mathcal{L}^1_\delta c_p\right]_{J-1}^2 \right) \sum_{\ell=1}^{J-1} \left(\sum_{k=0}^{n-1} (1 + \delta t \lambda_\ell)^{n-1-k} \mathrm{e}^{-p^2 \frac{\pi^2}{L^2} k \delta t} \right)^2 \\
&\quad + 2 \frac{\delta t^2}{J^2} \left(\left[\mathcal{L}^1_\delta c_p\right]_0^2 + \left[\mathcal{L}^1_\delta c_p\right]_{J-1}^2 \right) \left(\sum_{k=0}^{n-1} \mathrm{e}^{-p^2 \frac{\pi^2}{L^2} k \delta t} \right)^2.
\end{aligned} \tag{28}$$

For the first term in the estimate above, we observe that for all $p \geq 1$,

$$
\begin{aligned}
&\sum_{\ell=1}^{J-1}\left(\sum_{k=0}^{n-1}(1+\delta t\lambda_\ell)^{n-1-k}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}k\delta t}\right)^2 \\
&\quad = \sum_{\ell=1}^{J-1}\left(\sum_{k_1=0}^{n-1}(1+\delta t\lambda_\ell)^{n-1-k_1}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}k_1\delta t}\right)\left(\sum_{k_2=0}^{n-1}(1+\delta t\lambda_\ell)^{n-1-k_2}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}k_2\delta t}\right) \\
&\quad = \sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}(k_1+k_2)\delta t}\sum_{\ell=1}^{J-1}(1+\delta t\lambda_\ell)^{2n-2-k_1-k_2}.
\end{aligned}
$$

Using the CFL condition (13), we may use Proposition 3.2 to obtain

$$
\sum_{\ell=1}^{J-1}\left(\sum_{k=0}^{n-1}(1+\delta t\lambda_\ell)^{n-1-k}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}k\delta t}\right)^2 \leq \sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}(k_1+k_2)\delta t}\sum_{\ell=1}^{J-1}\mathrm{e}^{-\frac{\delta t}{\delta x^2}(2n-2-k_1-k_2)\sin^2\left(\frac{\ell\pi}{J}\right)}. \tag{29}
$$

One can split the sum on nonnegative numbers in $(k_1,k_2)\in\{0,\cdots,n-1\}^2$ above into the sum over the set $\mathcal{C}_n$ consisting in the $(k_1,k_2)\in\{0,\cdots,n-1\}^2$ with $2n-2-k_1-k_2\geq 1$ and the term where $k_1=k_2=n-1$. The latter is bounded by $(J-1)\mathrm{e}^{-p^2\frac{\pi^2}{L^2}(2n-2)\delta t}$. The former is bounded, thanks to Lemma 6.2, by

$$
\begin{aligned}
&\sum_{(k_1,k_2)\in\mathcal{C}_n}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}(k_1+k_2)\delta t}\sum_{\ell=1}^{J-1}\mathrm{e}^{-\frac{\delta t}{\delta x^2}(2n-2-k_1-k_2)\sin^2\left(\frac{\ell\pi}{J}\right)} \\
&\quad\leq \frac{1}{\delta x}\sum_{(k_1,k_2)\in\mathcal{C}_n}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}(k_1+k_2)\delta t}\delta x\sum_{\ell=1}^{J-1}\mathrm{e}^{-\frac{\delta t}{\delta x^2}(2n-2-k_1-k_2)\sin^2\left(\frac{\ell\pi}{J}\right)} \\
&\quad\leq \frac{1}{\delta x}\sum_{(k_1,k_2)\in\mathcal{C}_n}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}(k_1+k_2)\delta t}C\frac{\delta x}{\sqrt{(2n-2-k_1-k_2)\delta t}} \\
&\quad\leq C\sum_{(k_1,k_2)\in\mathcal{C}_n}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}(k_1+k_2)\delta t}\frac{1}{\sqrt{(2n-2-k_1-k_2)\delta t}}.
\end{aligned}
$$

where $C>1$ may depend on L but does not depend either on $n\geq 1$ or $k_1,k_2\in\{0,\cdots,n-1\}$ or $\delta t>0$ or $J\geq 2$ or $p\geq 1$ (see Lemma 6.2). Using this estimate in (29) and taking the term in $k_1=k_2=n-1$ into account, we infer for the term first term in the right-hand side of (28) above

$$
\begin{aligned}
&4\frac{\delta t^2}{J^2}\left(\left[\mathcal{L}_\delta^1 c_p\right]_0^2+\left[\mathcal{L}_\delta^1 c_p\right]_{J-1}^2\right)\sum_{\ell=1}^{J-1}\left(\sum_{k=0}^{n-1}(1+\delta t\lambda_\ell)^{n-1-k}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}k\delta t}\right)^2 \\
&\quad\leq 4\frac{\delta t^2}{J^2}\left(\left[\mathcal{L}_\delta^1 c_p\right]_0^2+\left[\mathcal{L}_\delta^1 c_p\right]_{J-1}^2\right)\left((J-1)\mathrm{e}^{-p^2\frac{\pi^2}{L^2}2(n-1)\delta t}+C\sum_{(k_1,k_2)\in\mathcal{C}_n}\frac{\mathrm{e}^{-p^2\frac{\pi^2}{L^2}(k_1+k_2)\delta t}}{\sqrt{(2n-2-k_1-k_2)\delta t}}\right) \\
&\quad\leq \frac{4C}{J^2}\left(\left[\mathcal{L}_\delta^1 c_p\right]_0^2+\left[\mathcal{L}_\delta^1 c_p\right]_{J-1}^2\right)\left((J-1)\delta t^2\mathrm{e}^{-p^2\frac{\pi^2}{L^2}2(n-1)\delta t}+\delta t^2\sum_{(k_1,k_2)\in\mathcal{C}_n}\frac{\mathrm{e}^{-p^2\frac{\pi^2}{L^2}(k_1+k_2)\delta t}}{\sqrt{(2n-2-k_1-k_2)\delta t}}\right),
\end{aligned}
$$

where we used $C>1$ and distributed δt^2 in the sum.

On the one hand, using (13), we have

$$(J-1)\delta t^2 \le \frac{\sqrt{2}}{2}(J-1)\delta x\,\delta t^{3/2} \le L\delta t^{3/2},$$

which is bounded since $\delta t \in (0,1)$. On the other hand, using Lemma 6.4, we infer that

$$\delta t^2 \sum_{(k_1,k_2)\in\mathcal{C}_n} \frac{\mathrm{e}^{-p^2\frac{\pi^2}{L^2}(k_1+k_2)\delta t}}{\sqrt{(2n-2-k_1-k_2)\delta t}} \le C,$$

for some $C>0$ that may depend on L but that does not depend on $n\ge 2$, $p\ge 1$, $\delta t\in(0,1)$. This implies

$$4\frac{\delta t^2}{J^2}\left(\left[\mathcal{L}_\delta^1 c_p\right]_0^2 + \left[\mathcal{L}_\delta^1 c_p\right]_{J-1}^2\right)\sum_{\ell=1}^{J-1}\left(\sum_{k=0}^{n-1}(1+\delta t\lambda_\ell)^{n-1-k}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}k\delta t}\right)^2 \le C\delta x^2\left(\left[\mathcal{L}_\delta^1 c_p\right]_0^2 + \left[\mathcal{L}_\delta^1 c_p\right]_{J-1}^2\right),$$

for some $C>0$ that may depend on L but that does not depend on $n\ge 2$, $p\ge 1$, $\delta t\in(0,1)$. This concludes the proof for the first term in the right-hand side of (28). Observe that a similar bound holds for the second term in the right-hand side of (28):

$$\begin{aligned}
&2\frac{\delta t^2}{J^2}\left(\left[\mathcal{L}_\delta^1 c_p\right]_0^2 + \left[\mathcal{L}_\delta^1 c_p\right]_{J-1}^2\right)\left(\sum_{k=0}^{n-1}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}k\delta t}\right)^2 \\
&\le \frac{2}{J^2}\left(\left[\mathcal{L}_\delta^1 c_p\right]_0^2 + \left[\mathcal{L}_\delta^1 c_p\right]_{J-1}^2\right)\left(\delta t\sum_{k=0}^{n-1}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}k\delta t}\right)^2 \\
&\le C\delta x^2\left(\left[\mathcal{L}_\delta^1 c_p\right]_0^2 + \left[\mathcal{L}_\delta^1 c_p\right]_{J-1}^2\right),
\end{aligned}$$

for some $C>0$ that may depend on L but that does not depend on $p\ge 1$, $n\ge 2$ or $\delta t\in(0,1)$ (see the proof of Lemma 6.4 for details). Using these estimates in (28) and taking square roots, we infer for all $p\ge 1$,

$$\left\|\delta t\sum_{k=0}^{n-1}(\mathsf{Id}+\delta t\mathsf{P}_\delta)^{n-1-k}\mathcal{L}_\delta^1 c_p(\cdot)\mathrm{e}^{-p^2\frac{\pi^2}{L^2}k\delta t}\right\|_{\ell^2} \le C\delta x\left(\left|\left[\mathcal{L}_\delta^1 c_p\right]_0\right| + \left|\left[\mathcal{L}_\delta^1 c_p\right]_{J-1}\right|\right).$$

Moreover, using (18), there exists $C>0$ such that for all $p\ge 1$,

$$\left|\left[\mathcal{L}_\delta^1 c_p\right]_0\right| + \left|\left[\mathcal{L}_\delta^1 c_p\right]_{J-1}\right| \le Cp^3.$$

Therefore, using the last two estimates above and (26), we obtain (25). $\square$

Analysis of the term with $\mathcal{L}_\delta^2$

Proposition 3.6. *Assume $L>0$ is fixed. There exists $C>0$ such that for all $u^0\in H^6(0,L)$ satisfying (6), for all $\delta t\in(0,1)$ and $J\ge 2$ such that the CFL condition (13) holds, and all $n\ge 1$,*

$$\left\|\delta t\sum_{k=0}^{n-1}(\mathsf{Id}+\delta t\mathsf{P}_\delta)^{n-1-k}\delta x^2\mathcal{L}_\delta^2 u(k\delta t,\cdot)\right\|_{\ell^2} \le C\delta x^2\sum_{p=1}^{+\infty}|\alpha_p|p^4. \tag{30}$$

Proof. Using the exact solution formula (3), we may write

$$\begin{aligned}\left\|\delta t\sum_{k=0}^{n-1}(\mathsf{Id}+\delta t\mathsf{P}_\delta)^{n-1-k}\delta x^2\mathcal{L}_\delta^2 u(k\delta t,\cdot)\right\|_{\ell^2} &\le \delta t\delta x^2\sum_{p=1}^{+\infty}|\alpha_p|\left\|\sum_{k=0}^{n-1}(\mathsf{Id}+\delta t\mathsf{P}_\delta)^{n-1-k}(\mathcal{L}_\delta^2 c_p)\mathrm{e}^{-p^2\frac{\pi^2}{L^2}k\delta t}\right\|_{\ell^2}\\ &\le \delta t\delta x^2\sum_{p=1}^{+\infty}|\alpha_p|\sum_{k=0}^{n-1}\left\|(\mathsf{Id}+\delta t\mathsf{P}_\delta)^{n-1-k}(\mathcal{L}_\delta^2 c_p)\right\|_{\ell^2}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}k\delta t}\\ &\le \delta t\delta x^2\sum_{p=1}^{+\infty}|\alpha_p|\left\|\mathcal{L}_\delta^2 c_p\right\|_{\ell^2}\sum_{k=0}^{n-1}\mathrm{e}^{-p^2\frac{\pi^2}{L^2}k\delta t},\end{aligned}$$

using also (14) because of the CFL condition (13). Since obviously, in view of (19),

$$\left\|\mathcal{L}_\delta^2 c_p\right\|_{\ell^2}\le Cp^4,$$

where $C>0$ may depend on L but does not depend on $p\ge 1$ or $J\ge 2$, we infer (30). □

Synthesis of the analysis of the term in $\mathcal{L}_\delta$

Plugging (25) and (30) into (17), we infer the following *uniform in time* error estimate for the term in ε_1 in (21).

Proposition 3.7. *Assume $L>0$ is fixed. There exists $C>0$ such that for all $u^0\in H^6(0,L)$ satisfying* (6)*, for all $\delta t\in(0,1)$ and $J\ge 2$ such that the CFL condition* (13) *holds, and all $n\ge 1$,*

$$\left\|\delta t\sum_{k=0}^{n-1}(\mathsf{Id}+\delta t\mathsf{P}_\delta)^{n-1-k}\mathcal{L}_\delta u(k\delta t,\cdot)\right\|_{\ell^2}\le C\delta x\sum_{p=1}^{+\infty}|\alpha_p|p^4. \tag{31}$$

3.3 Analysis of the term in ε_2

In contrast to the *uniform in time* error estimate (31) for the term in ε_1 in (21) obtained in Section 3.2, the following *uniform in time* error estimate (32) for the term in ε_2 in (21) is more simply obtained, using essentially the exponential decay of the x-derivatives of the exact solution of (1).

Proposition 3.8. *Assume $L>0$ is fixed. There exists $C>0$ such that for all $u^0\in H^6(0,L)$ satisfying* (6)*, for all $\delta t\in(0,1)$, $J\ge 2$, such that the CFL condition* (13) *holds, all $n\ge 1$,*

$$\left\|\sum_{k=0}^{n-1}(\mathsf{Id}+\delta t\mathsf{P}_\delta)^{n-1-k}\,\varepsilon_2^k\right\|_{\ell^2}\le C\delta t\|P^2u^0\|_{H^1}. \tag{32}$$

Proof. Since $u^0\in H^6(0,L)$, we have for all $t\ge 0$, $u(t)\in H^6(0,L)$. In particular, for all $t\ge 0$, $P^2u(t)\in H^2(0,L)$. Using Lemma 6.5, we infer that for all $t\ge 0$,

$$\|\Pi_{\delta x}P^2u(t)\|_{\ell^2}\le C\|P^2u(t)\|_{H^1}.$$

Using the exponential decay properties (7) and (8) of Property 2.6, we have

$$\forall t\ge 0,\qquad \|P^2u(t)\|_{H^1}\le \mathrm{e}^{-\frac{\pi^2}{L^2}t}\|P^2u^0\|_{H^1}.$$

This implies that

$$\forall t\ge 0,\qquad \|\Pi_{\delta x}P^2u(t)\|_{\ell^2}\le C\mathrm{e}^{-\frac{\pi^2}{L^2}t}\|P^2u^0\|_{H^1}.$$

The condition (13) implies that (14) holds. This allows to obtain

$$
\begin{aligned}
\left\| \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} \varepsilon_2^k \right\|_{\ell^2} &\le \sum_{k=0}^{n-1} \left\| (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} \varepsilon_2^k \right\|_{\ell^2} \\
&\le \sum_{k=0}^{n-1} \left\| \varepsilon_2^k \right\|_{\ell^2} \\
&\le \delta t^2 \sum_{k=0}^{n-1} \left\| \int_0^1 (1-\sigma) \Pi_{\delta x} P^2 u(k\delta t + \sigma \delta t) \mathrm{d}\sigma \right\|_{\ell^2} \\
&\le \delta t^2 \sum_{k=0}^{n-1} \int_0^1 (1-\sigma) \left\| \Pi_{\delta x} P^2 u(k\delta t + \sigma \delta t) \right\|_{\ell^2} \mathrm{d}\sigma \\
&\le C \delta t^2 \sum_{k=0}^{n-1} \int_0^1 (1-\sigma) \left\| P^2 u(k\delta t + \sigma \delta t) \right\|_{H^1} \mathrm{d}\sigma \\
&\le C \delta t^2 \sum_{k=0}^{n-1} \int_0^1 (1-\sigma) \mathrm{e}^{-\frac{\pi^2}{L^2}(k+\sigma)\delta t} \mathrm{d}\sigma \times \|P^2 u^0\|_{H^1} \\
&\le C \delta t^2 \sum_{k=0}^{n-1} \mathrm{e}^{-\frac{\pi^2}{L^2} k \delta t} \underbrace{\int_0^1 (1-\sigma) \mathrm{e}^{-\frac{\pi^2}{L^2} \sigma \delta t} \mathrm{d}\sigma}_{\le \frac{1}{2}} \times \|P^2 u^0\|_{H^1} \\
&\le C \frac{\delta t^2}{2} \sum_{k=0}^{n-1} \mathrm{e}^{-\frac{\pi^2}{L^2} k \delta t} \times \|P^2 u^0\|_{H^1} \\
&\le C \frac{\delta t^2}{2} \frac{1}{1 - \mathrm{e}^{-\frac{\pi^2}{L^2}\delta t}} \times \|P^2 u^0\|_{H^1} \\
&\le C \frac{L^2 \delta t}{2\pi^2} \times \frac{\frac{\pi^2}{L^2}\delta t}{1 - \mathrm{e}^{-\frac{\pi^2}{L^2}\delta t}} \times \|P^2 u^0\|_{H^1}.
\end{aligned}
$$

This proves (32) since the function $s \mapsto s/(1 - \mathrm{e}^{-s})$ is bounded over $(0, \pi^2/L^2)$. □

3.4 Synthesis of the analysis and proof of the main result

The convergence error e^n of the numerical scheme (12) applied to the linear heat equation (1) is split into 3 terms (see Equation (21)). For the first term in (21), for fixed $L > 0$, we have for all $J \ge 2$, $\delta t > 0$ such that the CFL condition (13) holds,

$$
\forall n \in \mathbb{N}, \qquad \| (\mathsf{Id} + \delta t \mathsf{P}_\delta)^n \, \mathsf{e}^0 \|_{\ell^2} \le \|\mathsf{e}^0\|_{\ell^2} = \|\Pi_{\delta x} u^0 - \mathsf{v}^0\|_{\ell^2}. \tag{33}
$$

For the second term in (21), we use Proposition 3.7 (hence Propositions 3.5 and 3.6) to obtain (31). For the third term in (21), we use Proposition 3.8 to obtain (32). This proves Theorem 2.12.

4 Application to the computation of a steady state

4.1 The continuous nonhomogeneous setting

As a by-product of the analysis carried out in Section 3 that led to the proof of Theorem 2.12, one obtains an explicit, albeit complicated, formula that solves the naively discretized stationary pure

Neumann problem (see (34) below) with order $\mathcal{O}(\delta x)$. This is consistent with the classical compensation of the inconsistency of a Dirichlet-Neumann problem by the (reinforced) stability property (see [15]).

For $f \in L^2(0,L)$, $\beta, \gamma \in \mathbb{R}$, the nonhomogeneous stationary Neumann problem reads

$$\begin{cases} -\partial_x^2 \tilde{u}^\infty &= f \\ \partial_x \tilde{u}^\infty(0) &= \beta \\ \partial_x \tilde{u}^\infty(L) &= \gamma. \end{cases} \tag{34}$$

If there exists a solution $\tilde{u}^\infty$ to (34), it is a steady state of

$$\begin{cases} \partial_t \tilde{u} - \partial_x^2 \tilde{u} &= f \\ \tilde{u}(0,x) &= \tilde{u}^0(x) \\ \partial_x \tilde{u}(t,0) &= \beta \\ \partial_x \tilde{u}(t,L) &= \gamma, \end{cases} \tag{35}$$

which is a nonhomogeneous version of (1).

Note that (34) is ill-posed. First, solutions to (34) may fail to exist in $H^2(0,L)$. Indeed, a necessary and sufficient compatibility condition for (34) to have a solution is

$$\gamma - \beta + \int_0^L f = 0. \tag{36}$$

Second, solutions to (34) are never unique: There is an additive degree of freedom since adding any constant function to a solution of (34) yields another solution.

Here is a physical interpretation of (36) in thermodynamics. Assume that $\tilde{u}(t,x)$ is the temperature in a metal rod of size L at time t and position x, and there is a steady heat source $f(x)$ and β and γ are the temperature gradients at $x = 0$ and $x = L$. The condition (36) means that in order for a steady state of (35) to exist, the total energy due to the source f must be balanced by the energy fluxes at the boundary. For example, if $\beta > 0$ and $\gamma = 0$, (36) means that the temperature gradient is positive at $x = 0$, so the heat exits the bar leftwards, thus the integral of the source f over the domain should be positive (and be equal to β), so the production of energy inside the bar compensates the energy that leaves the bar at the left end.

In contrast to (34), the problem (35) is well-posed regardless of (36). Nevertheless, assuming (36), the mean value of the solution $\tilde{u}$ of (35) satisfies $\partial_t \langle \tilde{u} \rangle = 0$. Moreover, still assuming (36), the solution $\tilde{u}$ of (35) converges exponentially fast in time to a solution $\tilde{u}^\infty$ to (34). Therefore, assuming (36), one can see (34) as the limit in time of Problem (35) and compute from $\tilde{u}^0$ the *unique*[1] solution $\tilde{u}^\infty \in H^2(0,L)$ of (34) with $\langle \tilde{u}^\infty \rangle = \langle \tilde{u}^0 \rangle$.

4.2 The discretized nonhomogeneous setting

The discretization of (34) is also ill-posed since the matrix P_δ is not invertible, and, even more dramatically, lacks consistency at both ends of the domain (see (17) of Proposition 2.11). However, the analysis of the convergence error that is conducted throughout Section 3 leads to writing a numerical scheme that, the degree of freedom $\langle \tilde{\mathsf{v}}^\infty \rangle_\delta$ being fixed, approximates the solution of (34), as the limit in time of the discrete time-dependent solution of

$$\tilde{\mathsf{v}}^{n+1} = (\mathsf{Id} + \delta t \mathsf{P}_\delta) \tilde{\mathsf{v}}^n + \delta t \, \mathsf{b}_{\delta x}, \tag{37}$$

[1] Using a classical result involving the Poincaré-Wirtinger inequality.

where

$$\mathsf{b}_{\delta x} = \Pi_{\delta x} f + (1/\delta x)(-\beta, 0, \ldots, 0, \gamma)^\top + r_{\delta x} 1, \tag{38}$$

for some small (with δx) real number $r_{\delta x}$ to be chosen later, and where $\tilde{\mathsf{v}}^0 \in \mathbb{R}^J$ is given. The system (37) is a discrete analogue to the system (35). Similarly, a discrete analogue to (34) reads

$$-\mathsf{P}_\delta \, \tilde{\mathsf{v}}^\infty = \mathsf{b}_{\delta x}, \tag{39}$$

where the definition of $\mathsf{b}_{\delta x}$ incorporates both a discretization of f and the boundary conditions of (34). In this setting, an analogue to the condition (36) is

$$\langle \mathsf{b}_{\delta x} \rangle_\delta = 0. \tag{40}$$

The condition (40), together with the relation (36) imposes the value of $r_{\delta x}$, namely

$$r_{\delta x} = \frac{1}{J\delta x} \left(\int_0^L f(x) \mathrm{d}x - \delta x \sum_{j=0}^{J-1} f(x_j) \right). \tag{41}$$

Observe that, provided f is smooth enough, we have $r_{\delta x} = \mathcal{O}(\delta x)$. The condition (40) means that $\mathsf{b}_{\delta x}$ belongs to the range of P_δ, which is the vector space $1^\perp$ orthogonal to 1. Of course, this is a necessary and sufficient compatibility condition for the existence of a solution to the non invertible linear system (39). Indeed, similarly to Problem (34), the system (39) is ill-posed, since $\mathrm{Ker}\ \mathsf{P}_\delta = \mathrm{Vect}(1)$.

As mentioned above, the analysis carried out in Section 3 ensures that, assuming (40) and the CFL condition (13), starting from $\tilde{\mathsf{v}}^0 \in \mathbb{R}^J$, the sequence $(\tilde{\mathsf{v}}^n)_{n\geq 0}$ defined by (37) converges exponentially fast to the unique solution $\tilde{\mathsf{v}}^\infty$ to (39) satisfying $\langle \tilde{\mathsf{v}}^\infty \rangle_\delta = \langle \tilde{\mathsf{v}}^0 \rangle_\delta$. Indeed, one has

$$\forall n \geq 0, \qquad \tilde{\mathsf{v}}^n = (\mathsf{Id} + \delta t \mathsf{P}_\delta)^n \tilde{\mathsf{v}}^0 + \delta t \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^k \, \mathsf{b}_{\delta x}. \tag{42}$$

Recall that $\mathsf{Id} + \delta t \mathsf{P}_\delta$ is diagonalizable with simple eigenvalues, and, under the CFL condition (13), 1 is an eigenvalue and the other eigenvalues are of modulus strictly less that 1. Consequently,

$$(\mathsf{Id} + \delta t \mathsf{P}_\delta)^n \tilde{\mathsf{v}}^0 \underset{n\to\infty}{\to} \langle \tilde{\mathsf{v}}^0 \rangle_\delta 1.$$

Observe that the power series with general term $(\mathsf{Id} + \delta t \mathsf{P}_\delta)^k$ does *not* converge in $\mathcal{M}_J(\mathbb{R})$. However, one has, using the fact that $\mathsf{b}_{\delta x}$ has zero mean value (see (40)) and the bound η of Proposition 2.8 using the assumption that the CFL condition (13) holds,

$$\forall k \geq 0, \qquad \|(\mathsf{Id} + \delta t \mathsf{P}_\delta)^k \, \mathsf{b}_{\delta x}\|_{\ell^2} \leq \eta^k \, \|\mathsf{b}_{\delta x}\|_{\ell^2}.$$

In particular, this implies that the sequence $\left(\sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^k \, \mathsf{b}_{\delta x}\right)_{n\geq 1}$ converges in $\mathbb{R}^J$. In conclusion, assuming (40) and (13), the only solution $\tilde{\mathsf{v}}^\infty$ of (39) with mean value $\langle \tilde{\mathsf{v}}^0 \rangle_\delta$ is

$$\tilde{\mathsf{v}}^\infty = \langle \tilde{\mathsf{v}}^0 \rangle_\delta 1 + \delta t \sum_{k=0}^{+\infty} \left[(\mathsf{Id} + \delta t \mathsf{P}_\delta)^k \, \mathsf{b}_{\delta x} \right]. \tag{43}$$

Observe that, despite the expression (43), the vector $\tilde{\mathsf{v}}^\infty$ does *not* depend on δt.

4.3 An interpretation of the previous result

Under the hypotheses (40) and (13), the convergence of the sequence $(\tilde{\mathsf{v}}^n)_{n\geq 0}$ generated by (37) towards the solution $\tilde{\mathsf{v}}^\infty$ to (39) with $\langle \tilde{\mathsf{v}}^\infty\rangle_\delta = \langle \tilde{\mathsf{v}}^0\rangle_\delta$ allows to interpret the explicit Euler scheme (37) for (35) as an iterative method of a (very simple) relaxation type for the ill-posed linear system (34): we rewrite (39) with $\mathsf{P}_\delta = \mathsf{M} - \mathsf{N}$ where $\mathsf{M} = (1/\delta t)(-\mathsf{Id})$ and $\mathsf{N} = -(1/\delta t)(\mathsf{Id} + \delta t\mathsf{P}_\delta)$. The matrix M is clearly invertible, and one has

$$\forall n \geq 0, \qquad \tilde{\mathsf{v}}^{n+1} = \mathsf{M}^{-1}\mathsf{N}\tilde{\mathsf{v}}^n + \mathsf{M}^{-1}\mathsf{b}_{\delta x}.$$

Observe that the spectral radius of $\mathsf{M}^{-1}\mathsf{N}$ is 1 so that the classical results for iterative relaxation methods do not apply. Nevertheless, thanks to (40), the sequence $(\tilde{\mathsf{v}}^n)_{n\geq 0}$ takes values in the affine space $\langle \tilde{\mathsf{v}}^0\rangle_\delta 1 + 1^\perp$, and the spectral radius of $\mathsf{M}^{-1}\mathsf{N}$ restricted to the stable subspace $1^\perp$ is the η defined in (15) which satisfies $0 \leq \eta < 1$.

4.4 Error estimates for the computation of the steady state

Similarly to the analysis of the error for the homogeneous case carried out in Section 3, we can analyse the error at time step number n defined by

$$\tilde{\mathsf{e}}^n = \Pi_{\delta x}\tilde{u}(\delta t) - \tilde{\mathsf{v}}^n,$$

by writing

$$\begin{aligned}
&\tilde{\mathsf{e}}^{n+1}\\
&= \Pi_{\delta x}\tilde{u}((n+1)\delta t) - \tilde{\mathsf{v}}^{n+1}\\
&= \Pi_{\delta x}\left(\tilde{u}(n\delta t) + \delta t(P\tilde{u}(n\delta t) + f) + \int_{n\delta t}^{(n+1)\delta t}((n+1)\delta t - s)\partial_t^2\tilde{u}(s)\mathrm{d}s\right) - (\mathsf{Id} + \delta t\mathsf{P}_\delta)\,\tilde{\mathsf{v}}^n - \delta t\mathsf{b}_{\delta x}\\
&= \Pi_{\delta x}\tilde{u}(n\delta t) + \delta t\Pi_{\delta x}P\tilde{u}(n\delta t) - (\mathsf{Id} + \delta t\mathsf{P}_\delta)\,\tilde{\mathsf{v}}^n + \delta t(\Pi_{\delta x}f - \mathsf{b}_{\delta x}) + \Pi_{\delta x}\int_{n\delta t}^{(n+1)\delta t}((n+1)\delta t - s)\partial_t^2\tilde{u}(s)\mathrm{d}s\\
&= (\mathsf{Id} + \delta t\mathsf{P}_\delta)\,\tilde{\mathsf{e}}^n + \underbrace{\delta t\left[(\Pi_{\delta x}P - \mathsf{P}_\delta\Pi_{\delta x})\,\tilde{u}(n\delta t) + (\Pi_{\delta x}f - \mathsf{b}_{\delta x})\right]}_{:=\tilde{\varepsilon}_1^n} + \underbrace{\int_{n\delta t}^{(n+1)\delta t}((n+1)\delta t - s)\Pi_{\delta x}\partial_t^2\tilde{u}(s)\mathrm{d}s}_{:=\tilde{\varepsilon}_2^n}.
\end{aligned}$$

This yields for all $n \geq 0$,

$$\tilde{\mathsf{e}}^n = (\mathsf{Id} + \delta t\mathsf{P}_\delta)^n\,\tilde{\mathsf{e}}^0 + \sum_{k=0}^{n-1}(\mathsf{Id} + \delta t\mathsf{P}_\delta)^{n-1-k}\,\tilde{\varepsilon}_1^k + \sum_{k=1}^{n-1}(\mathsf{Id} + \delta t\mathsf{P}_\delta)^{n-1-k}\,\tilde{\varepsilon}_2^k. \tag{44}$$

For the convergence of the numerical method (37) to a steady state of the nonhomogeneous heat equation (35), we prove the uniform result below.

Theorem 4.1. *Assume $L > 0$ is fixed. There exists $C > 0$ such that for all $f \in \mathcal{C}^3([0,L])$, $\beta,\gamma \in \mathbb{R}$ given such that (36) holds, all $\tilde{u}^0 \in H^6(0,L)$ such that*

$$\begin{cases}
\partial_x\tilde{u}^0(0) = \beta & \textit{and} \quad \partial_x\tilde{u}^0(L) = \gamma\\
\partial_x^3\tilde{u}^0(0) = -\partial_x f(0) & \textit{and} \quad \partial_x^3\tilde{u}^0(L) = -\partial_x f(L)\,,\\
\partial_x^5\tilde{u}^0(0) = -\partial_x^3 f(0) & \textit{and} \quad \partial_x^5\tilde{u}^0(L) = -\partial_x^3 f(L)
\end{cases} \tag{45}$$

all $\delta t \in (0,1)$ and $J \geq 2$ such that (13) *holds, all $n \in \mathbb{N}$, all $\tilde{\mathsf{v}}^0 \in \mathbb{R}^J$, one has*

$$\|\Pi_{\delta x}\tilde{u}(n\delta t) - \tilde{\mathsf{v}}^n\|_{\ell^2} \leq \left\|\Pi_{\delta x}\tilde{u}^0 - \tilde{\mathsf{v}}^0\right\|_{\ell^2} + C(\delta x + \delta t)\left(\|\tilde{u}^0\|_{H^5} + \|f\|_{H^3}\right), \tag{46}$$

where $\tilde{u}$ is the solution to (35) *and $(\tilde{\mathsf{v}}^n)_{n\geq 0}$ is the corresponding solution to* (37) *with $\mathsf{b}_{\delta x}$ defined in* (38)*-*(41)*.*

Remark 4.2. *Theorem 4.1 shows that, under the CFL condition* (13)*, the numerical method* (37) *is of order* 1 *in time and space for the computation of the stationary state of the nonhomogeneous heat equation* (35) *uniformly in time. Observe that starting numerically from $\tilde{\mathsf{v}}^0 = \Pi_{\delta x}\tilde{u}^0$ makes the first term in the right-hand side of* (46) *vanish.*

Remark 4.3. *The conditions* (45) *on the initial datum $\tilde{u}^0$ ensure that the function difference $u = \tilde{u} - \tilde{u}^\infty$ between the exact solution $\tilde{u}$ of the evolution nonhomogeneous heat equation* (35) *and the exact solution $\tilde{u}^\infty$ of the steady state equation* (34) *with $\langle \tilde{u}^\infty \rangle = \langle \tilde{u}^0 \rangle$ has a zero mean initial value $u(0) = \tilde{u}^0 - \tilde{u}^\infty$ satisfying* (6)*. In particular, the results of Section 2.2 apply to u.*

Note that this hypothesis (45) *allows to obtain an explicit and uniform in time bound, as described in the right-hand side of* (46)*. However, the relations* (45) *are probably not mandatory to ensure the uniform in time order of the method (see numerical experiments in Section 5). This remark is similar to Remark 2.4 in the homogeneous setting.*

Remark 4.4. *Observe that the estimate in the right-hand side of* (46) *is uniform in time. In particular, it does not depend on n. Moreover, we have $\tilde{u}(t) \underset{t\to+\infty}{\longrightarrow} \tilde{u}^\infty$ in $H^1(0,L)$ (Remark 4.3 implies that the difference between these two functions tends to* 0 *exponentially fast). Using Lemma 6.5, we infer that $\Pi_{\delta x}\tilde{u}(t) \underset{t\to+\infty}{\longrightarrow} \Pi_{\delta x}\tilde{u}^\infty$ in $\mathbb{R}^J$. Since $\tilde{\mathsf{v}}^n \underset{n\to+\infty}{\longrightarrow} \tilde{\mathsf{v}}^\infty$ in $\mathbb{R}^J$ (see* (43)*), we can pass to the limit in* (46) *to obtain*

$$\|\Pi_{\delta x}\tilde{u}^\infty - \tilde{\mathsf{v}}^\infty\|_{\ell^2} \leq \left\|\Pi_{\delta x}\tilde{u}^0 - \tilde{\mathsf{v}}^0\right\|_{\ell^2} + C(\delta x + \delta t)\left(\|\tilde{u}^0\|_{H^5} + \|f\|_{H^3}\right).$$

In particular, starting from $\tilde{\mathsf{v}}^0$ such that $\|\Pi_{\delta x}\tilde{u}^0 - \tilde{\mathsf{v}}^0\|_{\ell^2} = \mathcal{O}(\delta x)$, we infer that $\|\Pi_{\delta x}\tilde{u}^\infty - \tilde{\mathsf{v}}^\infty\|_{\ell^2} = \mathcal{O}(\delta x)$, so that the scheme (37) *computes an approximation of the steady state $\tilde{u}^\infty$ in $\mathcal{O}(\delta x)$.*

Proof. Recall that $\tilde{u}^\infty$ was defined at the end of Section 4.1 as *the* solution to (34) with $\langle \tilde{u}^\infty \rangle = \langle \tilde{u}^0 \rangle$. Proving Theorem 4.1 is done by estimating separately the three terms in the right-hand side of (44). Observe that

- Thanks to the CFL condition (13) and the stability property (14), the first term in the right-hand side of (44) is bounded easily and yields the first term in the right-hand side of (46).
- The third (and last) term in the right-hand side of (44) is easily bounded as in Proposition 3.8, by observing that the function $u(t) := \tilde{u}(t) - \tilde{u}^\infty$ is the solution to a linear homogeneous heat equation with homogeneous Neumann boundary conditions, with initial value $u(0) = \tilde{u}^0 - \tilde{u}^\infty$ with $\langle u(0) \rangle = 0$. Moreover, the functions $v = \partial_t u$ and $w = \partial_t v$, also solve the linear homogeneous heat equation over $(0, L)$ with homogeneous Neumann boundary conditions. Using (45), we infer that $u(0) = \tilde{u}^0 - \tilde{u}^\infty$ satisfies (6). In particular, one may use Proposition 2.6 to obtain that $w = \partial_t^2 u$ tends to 0 in H^1-norm exponentially fast, and so does $\partial_t^2 \tilde{u} = \partial_t^2 u$. A similar analysis as that of Proposition 3.8 therefore yields a term in $\delta t \times \|\tilde{u}^0\|_{H^5}$ in the right-hand side of (46).

Therefore, proving Theorem 4.1 amounts to finding a bound on the second term in the right-hand side of (44). Recall that the function u defined above is a solution of the homogeneous problem (1), with mean value 0, so that $\|u(t)\|_{H^1} \underset{t\to+\infty}{\longrightarrow} 0$, exponentially fast. Moreover, replacing $\tilde{u}(t)$ with $u(t)+\tilde{u}^\infty$, we have

$$\tilde{\varepsilon}_1^n = \delta t\,(\Pi_{\delta x}P - \mathsf{P}_\delta\Pi_{\delta x})\,u(n\delta t) + \delta t\,[(\Pi_{\delta x}P - \mathsf{P}_\delta\Pi_{\delta x})\,\tilde{u}^\infty + (\Pi_{\delta x}f - \mathsf{b}_{\delta x})]\,. \tag{47}$$

The analysis of the error for the first term in the expression above is the very same as in the homogeneous case carried out in Section 3.2, since u is a solution of the homogeneous problem (1) with zero mean value. This yields a term in δx in the right-hand side of (46), as it did in Proposition 3.7. Indeed, from Proposition 3.7, denoting by $(\alpha_p)_{p\geq 1}$ the coefficients of $u(0)=\tilde{u}^0-\tilde{u}^\infty$ in the cosine basis (4), we infer that

$$\begin{aligned}
\left\|\delta t\sum_{k=0}^{n-1}(I+\delta t\mathsf{P}_\delta)^{n-1-k}\mathcal{L}_\delta u(k\delta t)\right\| &\leq C\delta x\sum_{p=1}^{+\infty}|\alpha_p|p^5p^{-1}\\
&\leq C\delta x\left(\sum_{p=1}^{+\infty}\frac{1}{p^2}\right)^{\frac12}\left(\sum_{p=1}^{+\infty}\left(p^5|\alpha_p|\right)^2\right)^{\frac12}\\
&\leq C\delta x\|\partial_x^5u(0)\|_{L^2(0,L)}\\
&\leq C\delta x\left(\|\partial_x^5\tilde{u}^0\|_{L^2(0,L)}+\|\partial_x^5\tilde{u}^\infty\|_{L^2(0,L)}\right)\\
&\leq C\delta x\left(\|\tilde{u}^0\|_{H^5(0,L)}+\|\partial_x^3 f\|_{L^2(0,L)}\right),
\end{aligned}$$

where we have used (45) (which ensures that $u(0)$ satisfies (6)) to compute the coefficients of $\partial_x^5u(0)$ in the cosine basis of $L^2(0,L)$.

So, proving Theorem 4.1 amounts to proving an estimate for the part of the second term in (44) that corresponds to the last term in the right-hand side of (47). In the perspective of analysing this term, we may observe that, using the definition of $\mathsf{b}_{\delta x}$ in (38),

$$\mathcal{L}_\delta\tilde{u}^\infty + \Pi_{\delta x}f - \mathsf{b}_{\delta x} = \mathcal{L}_\delta\tilde{u}^\infty - (1/\delta x)(-\beta,0,\ldots,0,\gamma)^\top - r_{\delta x}1. \tag{48}$$

Using Definition 2.10 (of $\mathcal{L}_\delta$) and the fact that $\tilde{u}^\infty$ solves (34), we have

$$\langle\mathcal{L}_\delta\tilde{u}^\infty\rangle_\delta = \frac{1}{J}\sum_{j=0}^{J-1}(P\tilde{u}^\infty)(x_j) = -\frac{1}{J}\sum_{j=0}^{J-1}f(x_j).$$

This implies that

$$\langle\mathcal{L}_\delta\tilde{u}^\infty + \Pi_{\delta x}f - \mathsf{b}_{\delta x}\rangle_\delta = -\frac{1}{J}\sum_{j=0}^{J-1}f(x_j) + \frac{1}{J\delta x}(\beta-\gamma) - r_{\delta x}.$$

Using the definition of $r_{\delta x}$ in (41), we infer

$$\begin{aligned}
\langle\mathcal{L}_\delta\tilde{u}^\infty + \Pi_{\delta x}f - \mathsf{b}_{\delta x}\rangle_\delta &= -\frac{1}{J}\sum_{j=0}^{J-1}f(x_j) + \frac{1}{J\delta x}(\beta-\gamma) - \frac{1}{J\delta x}\int_0^L f(x)\mathrm{d}x + \frac{1}{J}\sum_{j=0}^{J-1}f(x_j)\\
&= \frac{1}{J\delta x}\left(\beta-\gamma-\int_0^L f(x)\mathrm{d}x\right).
\end{aligned}$$

Using the continuous condition (36), this proves that the second term in the error decomposition (47) has zero mean value. We observe that, using (38) and then the notation of Definition 2.10 and Proposition 2.11,

$$\begin{aligned} \mathcal{L}_\delta \tilde{u}^\infty + \Pi_{\delta x} f - \mathsf{b}_{\delta x} &= \mathcal{L}_\delta \tilde{u}^\infty + \Pi_{\delta x} f - \Pi_{\delta x} f - (1/\delta x)(-\beta, 0, \dots, 0, \gamma)^\top - r_{\delta x} 1 \\ &= \underbrace{\mathcal{L}^1_{\delta x} \tilde{u}^\infty}_{:=\Lambda_1} + \underbrace{\delta x^2 \mathcal{L}^2_{\delta x} \tilde{u}^\infty - r_{\delta x} 1}_{:=\Lambda_2}, \end{aligned}$$

since the boundary values of $\tilde{u}^\infty$ simplify. Since this vector $\Lambda_1 + \Lambda_2$ has zero mean value, we may write

$$\mathcal{L}_\delta \tilde{u}^\infty + \Pi_{\delta x} f - \mathsf{b}_{\delta x} = \Lambda_1 - \langle \Lambda_1 \rangle 1 + \Lambda_2 - \langle \Lambda_2 \rangle 1. \tag{49}$$

Denoting by $\Lambda_{1,0}$ and $\Lambda_{1,J-1}$ the first and last coefficients of the vector Λ_1 (the others vanish, according to Definition 2.10), we compute using Lemma 3.1 (see also (22) and (23))

$$\Lambda_1 - \langle \Lambda_1 \rangle 1 = \frac{\sqrt{2}}{J} \sum_{\ell=1}^{J-1} (\Lambda_{1,0} + (-1)^\ell \Lambda_{1,J-1}) \cos\left(\frac{\ell\pi}{2J}\right) \mathsf{W}_\ell.$$

Using the orthonormality of $(\mathsf{W}_\ell)_{1 \le \ell \le J-1}$, this implies that, for $J \ge 2$, $\delta t > 0$ such that (13) holds and $n \ge 1$, we have

$$\begin{aligned} \left\| \delta t \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} (\Lambda_1 - \langle \Lambda_1 \rangle 1) \right\|_{\ell^2}^2 &= \frac{2}{J^2} \sum_{\ell=1}^{J-1} \left| \delta t \sum_{k=0}^{n-1} (\Lambda_{1,0} + (-1)^\ell \Lambda_{1,J-1})(1 + \delta t \lambda_\ell)^k \cos\left(\frac{\ell\pi}{2J}\right) \right|^2 \\ &\le \frac{2}{J^2} \sum_{\ell=1}^{J-1} \left| \delta t \sum_{k=0}^{n-1} (|\Lambda_{1,0}| + |\Lambda_{1,J-1}|)(1 + \delta t \lambda_\ell)^k \right|^2 \\ &\le \frac{1}{J^2} (\Lambda_{1,0}^2 + \Lambda_{1,J-1}^2) \sum_{\ell=1}^{J-1} \left| \delta t \sum_{k=0}^{n-1} (1 + \delta t \lambda_\ell)^k \right|^2. \end{aligned}$$

Taking square roots and using Proposition 3.4, we infer that there exists a constant $C > 0$ such that for all such $J \ge 2$, $\delta t > 0$ such that the CFL condition (13) holds and all $n \ge 1$,

$$\left\| \delta t \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} (\Lambda_1 - \langle \Lambda_1 \rangle 1) \right\|_{\ell^2} \le C \delta x \max(|\Lambda_{1,0}|, |\Lambda_{1,J-1}|). \tag{50}$$

Observing (see (18)) that $2|\Lambda_{1,0}|$ and $2|\Lambda_{1,J-1}|$ are bounded by $\|(\tilde{u}^\infty)''\|_\infty + \delta x \|(\tilde{u}^\infty)^{(3)}\|_\infty$ yields a bound of the form δx times $(\|f\|_\infty + \|f'\|_\infty)$, which in turn is absorbed by $\delta x \|f\|_{H^3}$ in the right-hand side of (46).

The analysis of the second part of the decomposition (49) can be carried out using the triangle inequality, Proposition 3.2 and the definition of η in (15)

$$\left\| \delta t \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} (\Lambda_2 - \langle \Lambda_2 \rangle 1) \right\|_{\ell^2} \le \left(\delta t \sum_{k=0}^{n-1} \eta^k \right) \|\Lambda_2 - \langle \Lambda_2 \rangle 1\|_{\ell^2}.$$

On the one hand, using the CFL condition (13), we can apply Proposition 3.3, to obtain that

$$\left\| \delta t \sum_{k=0}^{n-1} (\mathsf{Id} + \delta t \mathsf{P}_\delta)^{n-1-k} (\Lambda_2 - \langle \Lambda_2 \rangle 1) \right\|_{\ell^2} \le 2L^2 \|\Lambda_2 - \langle \Lambda_2 \rangle 1\|_{\ell^2}. \tag{51}$$

On the other hand, in view of (19) and the definition (41) of $r_{\delta x}$, we have

$$\|\Lambda_2\|_{\ell^2} \leq \delta x^2 \|\mathcal{L}_\delta^2 \tilde{u}^\infty\|_{\ell^2} + |r_{\delta x}| \leq \frac{\delta x^2}{6} \|(\tilde{u}^\infty)^{(4)}\|_\infty + C\delta x \|f'\|_\infty.$$

And, similarly,

$$\|\langle \Lambda_2 \rangle 1\|_{\ell^2} \leq \frac{\delta x^2}{6} \|(\tilde{u}^\infty)^{(4)}\|_\infty + C\delta x \|f'\|_\infty.$$

Putting these two estimates in the right-hand side of (51) by triangle inequality, we obtain a bound in δx times $\|f'\|_\infty + \|f''\|_\infty$, which in turns is controlled by a term in δx times $\|f\|_{H^3}$ in the right-hand side of (46). This concludes the proof. □

As we shall see numerically in Section 5.2, the conclusion of Theorem 4.1 holds within it hypotheses, and also extends to weaker hypotheses. We will also illustrate in Section 5.3 how it extends to dimension 2 and allows to derive a new method for the computation of steady states of fully nonhomogeneous linear heat problems.

4.5 A remark on an alternative way to solve (34)

Another possible way to derive a numerical method to solve the nonhomogeneous linear heat equation with Neumann boundary conditions (34) is to consider the Laplace transform in time of the time-dependent heat equation (35). Of course, we assume in this section that the condition (36) holds. Setting for $s \in \mathbb{C}$ with $\mathrm{Re}\,(s) > 0$, $U(s,x) = \int_0^{+\infty} \mathrm{e}^{-st} \tilde{u}(t,x)\mathrm{d}t$, we obtain

$$\begin{cases} sU(s,x) - \partial_x^2 U(s,x) &= f(x) + \tilde{u}(0,x) \\ \partial_x U(s,0) &= \beta \\ \partial_x U(s,L) &= \gamma. \end{cases}$$

This motivates the introduction of the problem of finding $\tilde{u}_s^\infty$ as the solution to

$$\begin{cases} s\tilde{u}_s^\infty(x) - \partial_x^2 \tilde{u}_s^\infty(x) &= f(x) \\ \partial_x \tilde{u}_s^\infty(0) &= \beta \\ \partial_x \tilde{u}_s^\infty(L) &= \gamma, \end{cases} \tag{52}$$

Observe that this problem is well posed for all $f \in L^2(0,L)$, and that its solution $\tilde{u}_s^\infty$ has zero mean value. Then, one can check that $\tilde{u}_s^\infty$ converges to the solution $\tilde{u}^\infty$ with zero mean value when s tends to 0. Indeed, we have for all $s > 0$, $\alpha_{s,0} = 0$, $\alpha_0 = 0$, and for all $p \geq 1$,

$$\alpha_{s,p} = \frac{\gamma_p}{s + p^2\pi^2/L^2} \qquad \text{and} \qquad \alpha_p = \frac{\gamma_p}{p^2\pi^2/L^2},$$

where $\tilde{u}_s^\infty = \sum_{p=0}^{+\infty} \alpha_{s,p} c_p$, $\tilde{u}^\infty = \sum_{p=0}^{+\infty} \alpha_p c_p$, and $f = \sum_{p=0}^{+\infty} \gamma_p c_p$. This way, for $s > 0$,

$$\begin{aligned} \|\tilde{u}_s^\infty - \tilde{u}^\infty\|_{L^2}^2 &= \sum_{p=0}^{+\infty} (\alpha_{s,p} - \alpha_p)^2 = \sum_{p=1}^{+\infty} \left(\frac{1}{s + p^2\pi^2/L^2} - \frac{1}{p^2\pi^2/L^2} \right)^2 \gamma_p^2 \\ &= \frac{L^8}{\pi^8} s^2 \sum_{p=1}^{+\infty} \frac{1}{p^2 \left(\frac{L^2}{\pi^2} s + p \right)^2} \gamma_p^2 \leq \frac{L^8}{\pi^8} s^2 \|f\|_{L^2}^2. \end{aligned}$$

In particular, $\|\tilde{u}_s^\infty - \tilde{u}^\infty\|_{L^2} = \mathcal{O}(s)$ when s tends to 0, and the convergence follows. Observe that one has, similarly, $\|\tilde{u}_s^\infty - \tilde{u}^\infty\|_{H^1} = \mathcal{O}(s)$.

Similarly, in order to solve (39), we consider the problem of solving for $J \geq 2$ and $s > 0$

$$s\tilde{\mathsf{v}}_s^\infty - \mathsf{P}_\delta \tilde{\mathsf{v}}_s^\infty = \mathsf{b}_{\delta x}, \tag{53}$$

where $\mathsf{b}_{\delta x}$ is defined in (38) and satisfies (41). In this discrete setting, the system (53) is well-posed and its solution v_s^∞ satisfies $\langle \mathsf{v}_s^\infty \rangle_\delta = 0$. Moreover, denoting by v^∞ the solution to (34) with zero mean value, we have $\|\tilde{\mathsf{v}}_s^\infty - \tilde{\mathsf{v}}^\infty\|_{\ell^2} = \mathcal{O}(s)$ when s tends to 0 with $J \geq 2$ fixed. Observe that, in addition, the constant in this $\mathcal{O}$ does not depend on J. Indeed, denoting $\hat{\mathsf{x}}(\ell) := \langle \mathsf{x}, \mathsf{W}_\ell \rangle_\delta$ for any vector $\mathsf{x} \in \mathbb{R}^J$, we have, for $s > 0$, and $1 \leq \ell \leq J-1$,

$$\widehat{\tilde{\mathsf{v}}^\infty}(\ell) - \widehat{\tilde{\mathsf{v}}^\infty}_s(\ell) = \left[\frac{1}{-\lambda_\ell} - \frac{1}{s - \lambda_\ell}\right] \widehat{\mathsf{b}_{\delta x}}(\ell).$$

Since

$$\min_{1 \leq l \leq J-1} |\lambda_\ell| = \frac{4}{\delta x^2} \min_{1 \leq l \leq J-1} \sin^2\left(\frac{\pi \ell}{2J}\right) = \frac{4}{\delta x^2} \sin^2\left(\frac{\pi}{2J}\right) \leq \pi^2 \frac{(J-1)^2}{L^2 J^2} \leq \frac{\pi^2}{L^2},$$

we infer that, for $s > 0$,

$$\begin{aligned}
\|\tilde{\mathsf{v}}^\infty - \tilde{\mathsf{v}}_s^\infty\|_{\ell^2}^2 &= \sum_{\ell=1}^{J-1} \left(\frac{s}{-\lambda_\ell(s - \lambda_\ell)}\right)^2 |\widehat{\mathsf{b}_{\delta x}}(\ell)|^2 \\
&\leq s^2 \sum_{\ell=1}^{J-1} \frac{1}{\lambda_\ell^4} |\widehat{\mathsf{b}_{\delta x}}(\ell)|^2 \leq \frac{s^2}{\min_\ell |\lambda_\ell|^4} \|\mathsf{b}_{\delta x}\|_{\ell^2} \\
&\leq s^2 \frac{L^8}{\pi^8} \|\mathsf{b}_{\delta x}\|_{\ell^2}^2,
\end{aligned}$$

and the fact that $\|\tilde{\mathsf{v}}_s^\infty - \tilde{\mathsf{v}}^\infty\|_{\ell^2} = \mathcal{O}(s)$ with a constant that does not depend on J is proved.

Therefore, one may solve the well-posed discrete elliptic problem (53) for small $s > 0$ and consider it as an approximate solution to the continuous ill-posed problem (34) with zero mean value. With the notation above, the error in this strategy is bounded in the following way :

$$\|\tilde{\mathsf{v}}_s^\infty - \Pi_{\delta x} \tilde{u}^\infty\|_{L^2} \leq \|\tilde{\mathsf{v}}_s^\infty - \tilde{\mathsf{v}}^\infty\|_{L^2} + \|\tilde{\mathsf{v}}^\infty - \Pi_{\delta x} \tilde{u}^\infty\|_{L^2}. \tag{54}$$

The first term in the right-hand side above is bounded by a constant that does not depend on $J \geq 2$ times s, as we explained above. The second term in the right-hand side of (54) is bounded by δx times a constant that does not depend on $s > 0$, as we noticed in Remark 4.4. As a conclusion, this (direct) alternative method to solve (34), as opposed to the (iterative) method described in (37), produces an error in $\mathcal{O}(s) + \mathcal{O}(\delta x)$, and requires solving a symmetric, sparse and well-posed linear system of size J, the condition number of which tends to $+\infty$ as s tends to 0.

5 Numerical results

In this section, we present numerical experiments illustrating the relevance and sharpness of Theorem 2.12 for the homogeneous linear heat equation and of Theorem 4.1 for the nonhomogeneous linear heat equation. In Section 5.1 (respectively 5.2), we investigate how the scheme (12) (resp. (37)) behaves in dimension 1 for several initial data, which allows to discuss the relevance of the hypotheses of Theorem 2.12 (resp. 4.1). In Section 5.3, we perform numerical experiments on an nonhomogeneous linear heat equation (similar to (12)) in dimension 2, using an extension of the scheme (37) to this context and we demonstrate numerically the fact that the *uniform in time* order estimate (analogue to (46)) of

Theorem 4.1 that allows to solve the corresponding time-dependent linear nonhomogeneous equation in order to compute an approximation of a steady state still holds in dimension 2.

The Matlab code we developed can be found at `https://github.com/paulinelafitte/codes_dl`.

5.1 1D numerical experiments in the homogeneous setting

We consider in this section the homogeneous linear heat equation (1) in one dimension with homogeneous Neumann boundary conditions. Our goal is to illustrate numerically the result stated in Theorem 2.12: The error of the numerical scheme (12), for different values of δt and δx (through J) under the CFL condition (13), at time $n\delta t$, is bounded by $\mathcal{O}(\delta x)$ where the constant in the $\mathcal{O}$ can be chosen *independently* of n. We also aim at testing the relevance of the hypotheses of that Theorem. We perform three numerical experiments corresponding to three different initial data (a trigonometric polynomial that satisfies the hypothesis (6), a smooth function that does not satisfy the hypothesis (6), and a compactly supported function that does not either satisfy the hypothesis (6)). We comment them below.

First, we consider the trigonometric polynomial initial datum u^0 with $(\alpha_i)_{0\leq i\leq 5} = (1, 1, 5, -1, 2, 1)$ and $\alpha_i = 0$ for $i \geq 6$ where $(\alpha_i)_{i\in\mathbb{N}}$ are defined just after (4). We set the numerical initial datum $\mathsf{v}^0 = \Pi_{\delta x} u^0$ and the length of the interval $L = 1$. For several values of J, we set $\delta t = \delta x^2/2$ so that (13) is fulfilled. We plot in the left panel of Figure 1 the error $\|\Pi_{\delta x} u(n\delta t) - \mathsf{v}^n\|_{\ell^2}$ at final time $n\delta t$ for several values of $n\delta t$ as a function of $J \geq 1$ in logarithmic scale. We observe that, for large $J \geq 1$, the error is indeed bounded by a constant times δx that can be chosen independently of $n\delta t$. This illustrates the result of Theorem 2.12.

Next, we consider the initial datum

$$u_0 : \begin{pmatrix} (0, L) & \longrightarrow & \mathbb{R} \\ x & \longmapsto & x^2(L-x)^2 \end{pmatrix}. \tag{55}$$

This function satisfies $u_0 \in \mathrm{Dom}(P)$. However, $Pu_0 \notin \mathrm{Dom}(P)$. Therefore, u_0 does not satisfy (6). We set the numerical initial datum $\mathsf{v}^0 = \Pi_{\delta x} u^0$ and the length of the interval is still $L = 1$. For several values of J, we set $\delta t = \delta x^2/2$ so that (13) is fulfilled. We plot in the right panel of Figure 1 the error $\|\Pi_{\delta x} u(n\delta t) - \mathsf{v}^n\|_{\ell^2}$ at final time $n\delta t$ for several values of $n\delta t$ as a function of $J \geq 1$ in logarithmic scale. We observe that, for large $J \geq 1$, the error is once again bounded by a constant times δx that can be chosen independently of $n\delta t$. This indicates that the result of Theorem 2.12 seems to hold even if u_0 does not satisfy all the hypotheses of the theorem.

Last, we consider $L = 2$ and the initial datum

$$u_0 : \left(\begin{matrix} (0, L) & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \max\left(1 - \frac{|L/2-x|}{\ell}, 0\right) \end{matrix} \right), \tag{56}$$

with $\ell = 1/50$. This initial datum is compactly supported in $(0, L)$. Even though it does not satisfy the hypothesis (6), we may expect Theorem 2.12 to hold, as illustrated numerically and explained above. The numerical initial datum is set to $\mathsf{v}^0 = \Pi_{\delta x} u_0$. The numerical results are displayed in Figure 2. The diffusion equation (1) takes some time to extend the support of the initial datum significantly up to the boundary of the domain. Therefore, in the early times of the dynamics (for $t = n\delta t < 0.02$), the numerical analysis carried out to prove Theorem 2.12 goes as if $\mathcal{L}_\delta^1 u(t)$ (see (18) in Proposition 2.11) was zero (its cumulated contribution in the term in ε_1 in (21) remains below the size of the other terms). For these short times ($t = n\delta t < 0.02$), the term in (25) plays no role in the error, and one

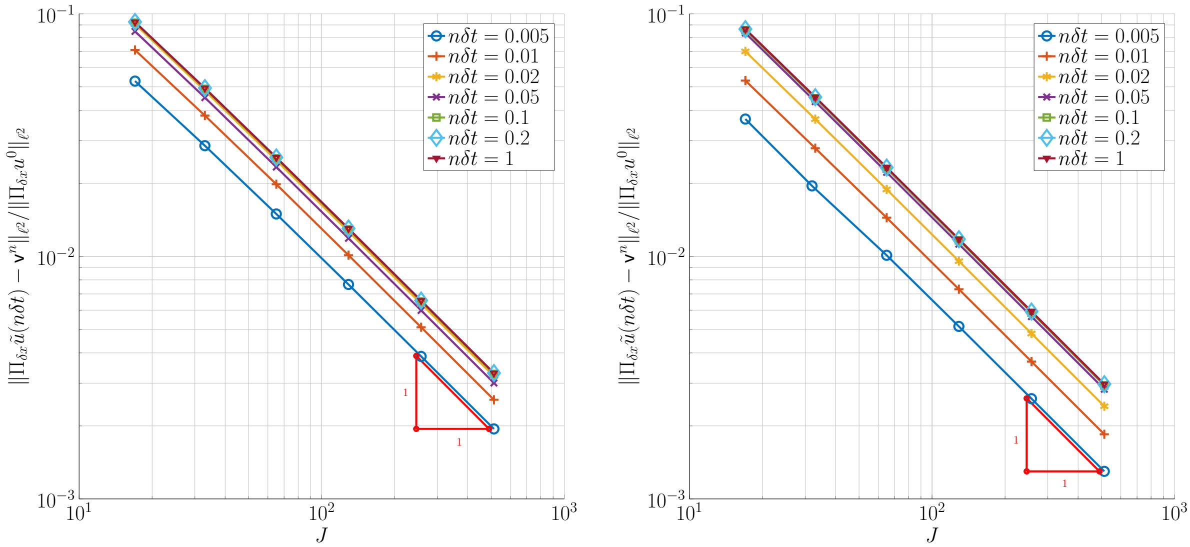

Figure 1: Numerical error as a function of J for several values of $n\delta t$ when approximating the solution of the homogeneous linear heat equation (1) using the scheme (12) under the CFL condition (13) (logarithmic scales). The initial datum given in the text : trigonometric polynomial (left panel) and function (55) (right panel).

only sees the term in (30) in ε_1. Therefore, the bound in (20) is in $\mathcal{O}(\delta x^2 + \delta t) = \mathcal{O}(\delta x^2)$ for not too large values of J. In contrast, for larger times ($n\delta t > 0.05$), the support of the exact solution reaches significantly the boundary of the domain and the term in $\mathcal{L}^1_\delta u(t)$ in ε_1 can no longer be neglected. For these times, and for all values of J, the bound (20) of Theorem 2.12 is this time in $\mathcal{O}(\delta x + \delta t) = \mathcal{O}(\delta x)$. This numerical result illustrates once again that the conclusion of Theorem 2.12 holds true even if all the hypotheses are not met, and that the first splitting of the error in (21) and the second splitting of the error in ε_1 using the decomposition of $\mathcal{L}_\delta$ in Proposition 2.11 are relevant for this numerical scheme applied to this problem : When the exact solution has a significant nonzero contribution at the boundary of the domain, it cannot be ignored in the analysis of the scheme, and it can anyway be controlled to prove a bound as (20) in Theorem 2.12.

5.2 1D numerical experiments in the nonhomogeneous setting

In the nonhomogeneous setting of Section 4, we consider a given source f and fluxes β and γ such that the balance equation (36) between the source and the fluxes is fulfilled. We compute approximations of the steady state $\tilde{u}^\infty$ solution to (34) (with a constraint on its mean value). To do so, we implement the algorithm (37), which produces approximations of the solution $\tilde{u}$ of the nonhomogeneous time-dependant heat equation (35), associated to some initial datum $\tilde{u}^0$ with the same mean value as $\tilde{u}^\infty$. Our goal is to illustrate numerically the validity of Theorem 4.1, and in particular of the bound (46), to discuss the necessity of its hypothesis, and to demonstrate how it allows to compute numerical approximations of the steady state $\tilde{u}^\infty$. We consider the case $L = 2$ and the continuous and piecewise affine source term

$$f : x \longmapsto \begin{cases} 1 & \text{if } 0 \le x \le 1/2, \\ 2x & \text{if } 1/2 < x \le 2, \end{cases}$$

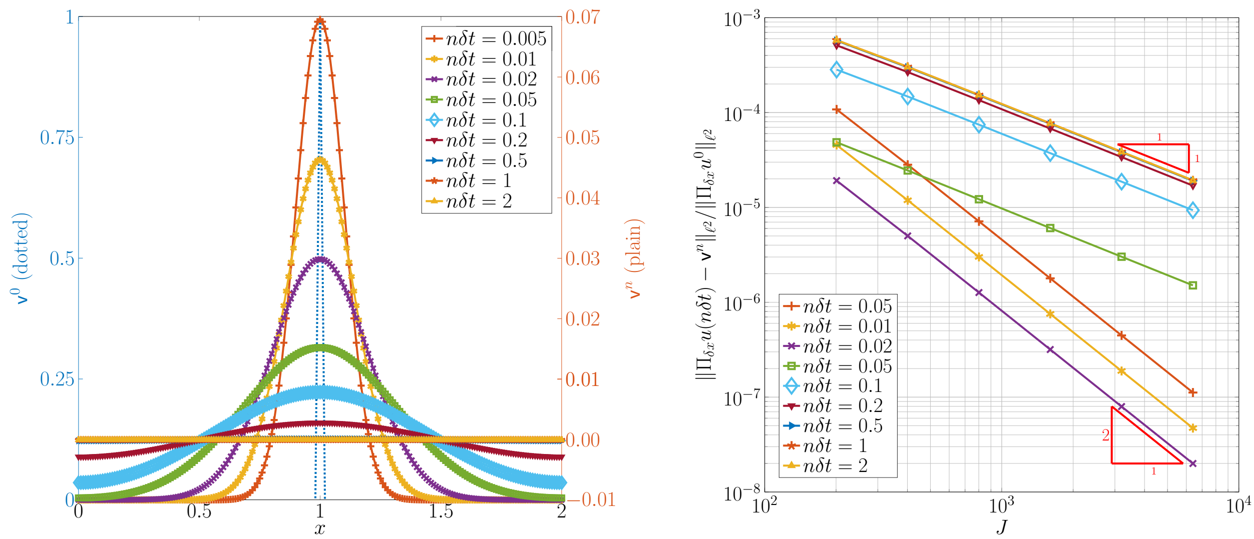

Figure 2: Numerical simulation of the solution of the homogeneous linear heat equation (1) associated to the initial datum (56) using the scheme (12) under the CFL condition (13) for several values of $n\delta t$. Left : Numerical approximation v^n of the solution $u(n\delta t)$ as a function of $j\delta x$ computed for several values of $n\delta t$ with $J = 201$ (multiple scales). Right : Numerical error as a function of J for the same values of $n\delta t$ (logarithmic scales).

and the boundary conditions $\beta = 1/2$ and $\gamma = -15/4$. Since

$$\int_0^L f(x)\mathrm{d}x = (1/2 + (2^2 - (1/2)^2)) = 17/4 = \beta - \gamma,$$

we have that (36) holds. The solution $\tilde{u}^\infty$ to (34) with $\langle \tilde{u}^\infty \rangle = -\frac{193}{384}$ reads

$$\tilde{u}^\infty : x \longmapsto \begin{cases} -\frac{(x-1/2)^2}{2} & \text{if } 0 \le x \le 1/2, \\ -\frac{x^3}{3} - \frac{1}{4}x - \frac{1}{12} & \text{if } 1/2 < x \le 2. \end{cases}$$

We define the function

$$w : x \longmapsto \frac{\gamma - \beta}{2L}x^2 + \beta x - \frac{\gamma - \beta}{6}L - \frac{\beta}{2}L,$$

which satisfies $-\partial_x^2 w = 0$ over $(0, L)$, the boundary conditions $\partial_x w(0) = \beta$ and $\partial_x w(L) = \gamma$ (which correspond to the first two lines of (45)), and has zero mean value.

We use two different initial data as $\tilde{u}^0$ and $\tilde{\mathsf{v}}^0$:

- First, we consider $\tilde{u}^0 = \langle \tilde{u}^\infty \rangle \mathbb{1} + w$, and $\tilde{\mathsf{v}}^0 = \Pi_{\delta x}\tilde{u}^0$. In particular, in this case, $\tilde{u}^0$ satisfies the first line of the hypotheses (45), and the first term in the right hand side of (46) vanishes.
- Second, we consider $\tilde{u}^0 = \langle \tilde{u}^\infty \rangle \mathbb{1}$, and $\tilde{\mathsf{v}}^0 = \Pi_{\delta x}\tilde{u}^0$. In particular, in this case, $\tilde{u}^0$ *does not* satisfy the first line of the hypotheses (45), and the first term in the right hand side of (46) still vanishes.

For the interpretation of the numerical results displayed in Figure 3, we point out the inequality, that is valid for all $J \ge 2$ and $\delta t > 0$ such that (13) holds and all $n \in \mathbb{N}$,

$$\|\Pi_{\delta x}\tilde{u}^\infty - \tilde{\mathsf{v}}^n\|_{\ell^2} \le \|\Pi_{\delta x}\tilde{u}^\infty - \Pi_{\delta x}\tilde{u}(n\delta t)\|_{\ell^2} + \|\Pi_{\delta x}\tilde{u}(n\delta t) - \tilde{\mathsf{v}}^n\|_{\ell^2}. \tag{57}$$

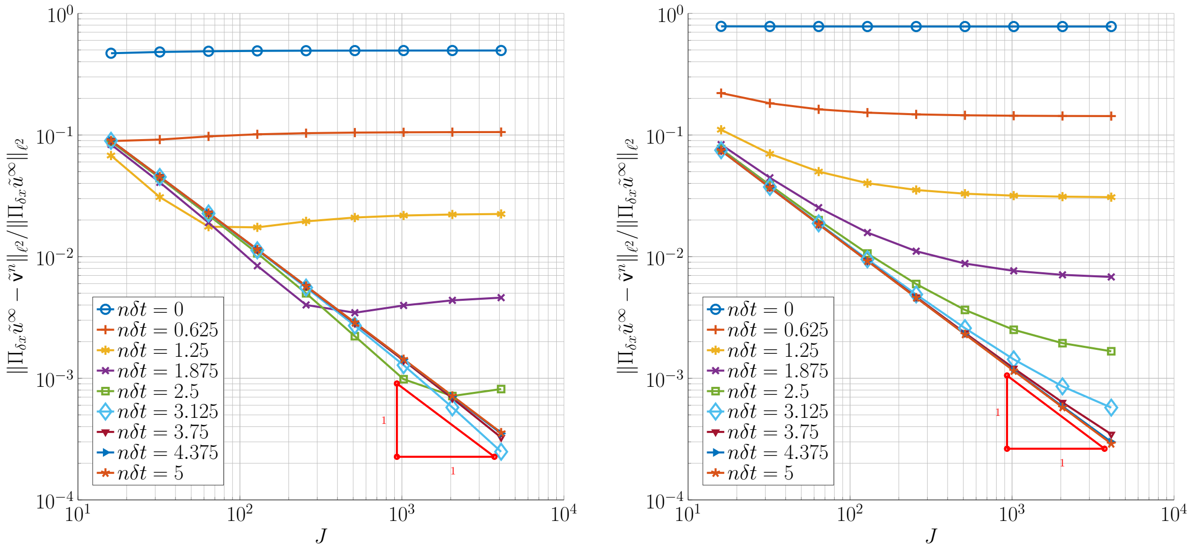

Figure 3: Numerical error as a function of J for several values of $n\delta t$ when approximating the solution of the linear nonhomogeneous stationary heat equation (34) using the scheme (37) under the CFL condition (13) (logarithmic scales). Initial datum : $\tilde{u}^0 = \langle \tilde{u}^\infty \rangle \mathbb{1} + w$ (left panel) and $\tilde{u}^0 = \langle \tilde{u}^\infty \rangle \mathbb{1}$ (right panel).

The error displayed in Figure 3 corresponds to the left hand side of this inequality. In the right-hand side of (57),

- the first term tends to 0 exponentially fast and independently of $J \geq 2$ when $n\delta t$ tends to $+\infty$ because $\tilde{u}$ solves (35), $\tilde{u}^\infty$ solves (34) and $\langle \tilde{u}^0 \rangle = \langle \tilde{u}^\infty \rangle$,
- thanks to (46) of Theorem 4.1, the second term is bounded *independently of* n, under the CFL condition (13) (note that the first term in the right-hand side of (46) vanishes in our two cases), by $\mathcal{O}(\delta x)$.

Therefore, for a fixed $n\delta t$ and varying δx, under the CFL condition (13), we should see (in logarithmic scales, when plotting the error as a function of $J \geq 2$) a straight line of slope -1 that starts to stall when the first term in the right hand side of (57) becomes bigger than the second term. Moreover, all these errors (no matter the value of $n\delta t$, in particular when it is big), in the regime when they are in $\mathcal{O}(\delta x)$ (straight lines of slope -1), remain under a common straight line of slope -1. This illustrates the fact that the constant C in Theorem 4.1 *does not* depend on $n\delta t$ nor $J \geq 2$ provided that the CFL condition (13) holds.

These numerical results illustrate that the conclusion of Theorem 4.1 holds true way beyond its hypotheses. First, observe that the source term f is *not* in $\mathcal{C}^3([0, L])$. Second, the two initial data described above *do not* satisfy the hypothesis (45). Indeed, the hypotheses (45) appear as technical hypotheses ensuring the simplicity of the proof in absence of initial layer in the solution of (35), so that $u(t) = \tilde{u}(t) - \tilde{u}^\infty$ is a solution of the linear homogeneous heat equation (1) with an initial datum satisfying (6), that ensures that $u(0)$, $Pu(0)$ and $P^2u(0)$ are in the domain of P (see also Remark 2.4 for the homogeneous setting). This point allows to consider virtually any initial datum in $L^2(0, L)$ with the correct mean value to compute numerically approximations of $\tilde{u}^\infty$ using the scheme (37). This is of particular importance since computing the function w above may not be accessible in higher dimensions and more complicated geometries, as is illustrated in the next Section.

5.3 2D numerical experiments in the nonhomogeneous setting

We consider, for some $\alpha, \beta > 0$ and $x_0, y_0 \in \mathbb{R}$,

- the rectangular domain $\Omega = (0,2) \times (0,4)$, and its boundary $\Gamma = \overline{\Omega} \setminus \Omega = \Gamma_1 \cup \Gamma_2$ where $\Gamma_1 = (\{0\} \times [0,4]) \cup (\{2\} \times [0,4]\})$ and $\Gamma_2 = ([0,2] \times \{0\}) \cup ([0,2] \times \{4\})$,
- the source term defined over Ω

$$f : (x,y) \longmapsto (2\alpha(1 - 2\alpha(x-x_0)^2) + 2\beta(1 - 2\beta(y-y_0)^2))\, \mathrm{e}^{-\alpha(x-x_0)^2 - \beta(y-y_0)^2},$$

- the boundary conditions $g_1 : (x,y) \longmapsto -2\alpha(x-x_0)\, \mathrm{e}^{-\alpha(x-x_0)^2-\beta(y-y_0)^2}$ defined over Γ_1 and $g_2 : (x,y) \longmapsto -2\beta(y-y_0)\, \mathrm{e}^{-\alpha(x-x_0)^2-\beta(y-y_0)^2}$ defined over Γ_2.

We aim at solving for $\tilde{u}^\infty \in H^2(\Omega)$ the nonhomogeneous stationary heat equation

$$-(\partial_x^2 + \partial_y^2)\tilde{u}^\infty = f, \qquad (x,y) \in \Omega, \tag{58}$$

with the nonhomogeneous Neumann boundary conditions

$$\begin{cases} \partial_x \tilde{u}^\infty &= g_1, \qquad (x,y) \in \Gamma_1, \\ \partial_y \tilde{u}^\infty &= g_2, \qquad (x,y) \in \Gamma_2. \end{cases} \tag{59}$$

The conditions (58)-(59) define $\tilde{u}^\infty$ in $H^2(\Omega)$ up to a constant. As in the one-dimensional case (see Section 4), this problem can be alleviated by imposing the mean value of $\tilde{u}^\infty$. The source term f and the heat fluxes g_1, g_2 are designed so that an exact solution of (58)-(59) is $\tilde{u}^\infty : (x,y) \longmapsto \mathrm{e}^{-\alpha(x-x_0)^2-\beta(y-y_0)^2}$. We use a $2d$-version of the scheme (37) to approximate this exact solution as if we did not know it, as we did in Section 5.2 in the one-dimensional case.

The numerical simulations given below were computed for the parameter $\beta = 5$, starting from $\tilde{\mathrm{v}}^0 = 0$. The Matlab code, that we wrote to obtain the following plots, was written starting from a resolution of a 2D Laplace equation with a mixed Dirichlet-Neumann boundary [17]. We run a first simulation with $\alpha = 15$, $x_0 = 1$ and $y_0 = 2$, so that the Neumann condition at the boundary is very small, compared to the numerical errors. We run another simulation with $\alpha = 1$, $x_0 = 0$ and $y_0 = 4$, so that the Neumann condition on the boundary can no longer be neglected. The numerical results are displayed in Figure 4. On the top panels of Figure 4, the effect of the nonzero Neumann boundary condition seems to be actually negligible : the numerical error seems to be bounded by $\mathcal{O}(\delta x^2)$ when $n\delta t$ is taken sufficially big. On the bottom panels of Figure 4, the effect of the nonzero Neumann boundary condition can no longer be neglected and the scheme behaves as it does in dimension 1 (see Section 5.2) in accordance with Theorem 4.1 : the numerical error seems to be bounded by $\mathcal{O}(\delta x^1)$ when $n\delta t$ is taken sufficially big. Note that, for $n\delta t = 5$ and $J = 32$, the error is still big enough for the maximum of $\tilde{\mathrm{v}}^n$ to be close to 1.5 (bottom left panel), while it was close to 1.0 in the centered case (top left panel).

6 Appendix

6.1 Proof of Proposition 3.4

Proof. Since for all $\ell \in \{1, \cdots, J-1\}$, $\lambda_\ell \neq 0$. For $\delta t > 0$, this implies that $1 + \delta t \lambda_\ell \neq 1$ for $\ell \in \{1, \cdots, J-1\}$. Assuming that $J \geq 2$ and $\delta t > 0$ satisfy the CFL condition (13), we have

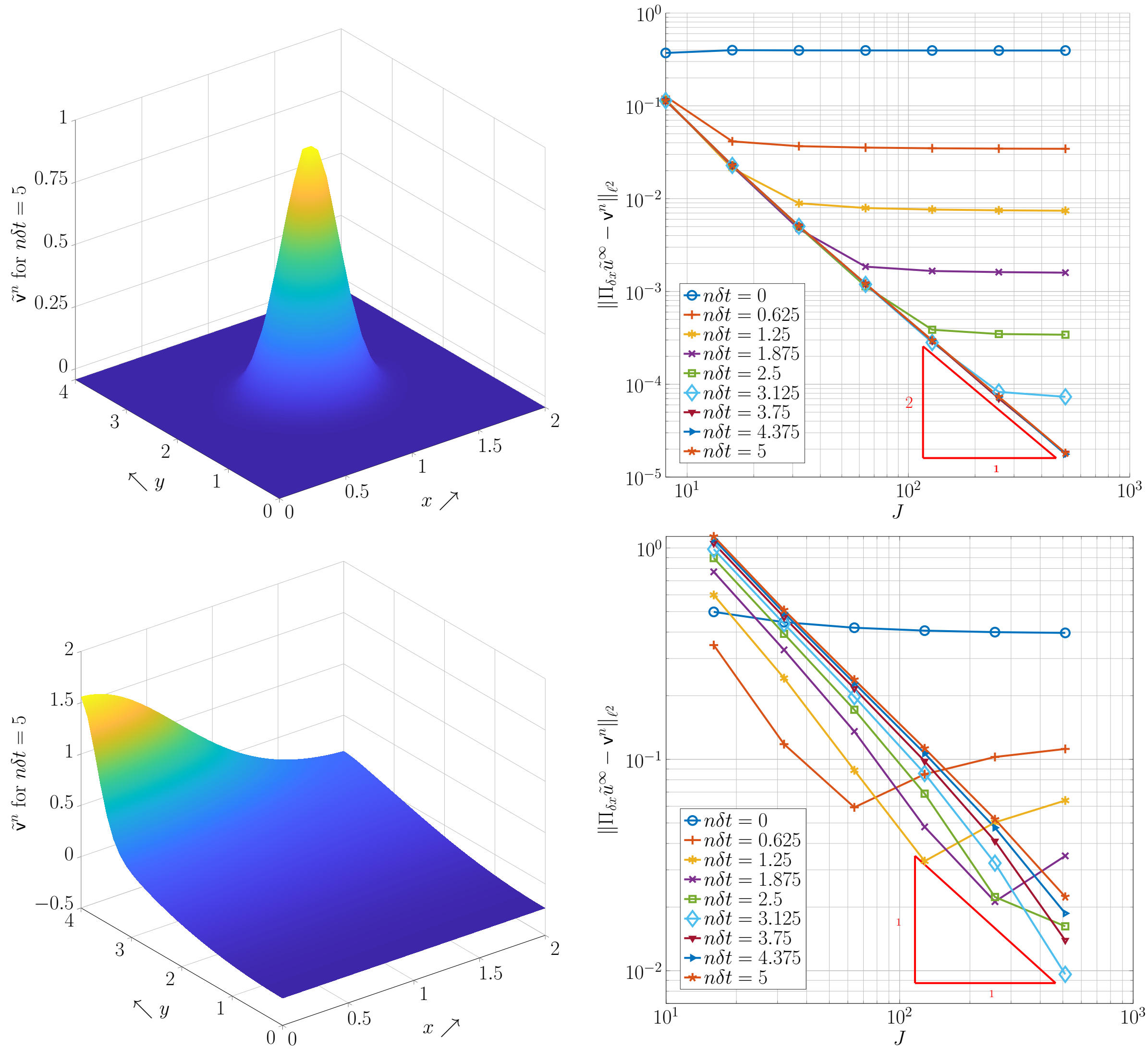

Figure 4: On the left panels, numerical solution $\tilde{\mathrm{v}}^n$ obtained at $n\delta t = 5$ and $J = 32$. On the right panels, numerical error $\|\Pi_{\delta x}\tilde{u}^\infty - \tilde{\mathrm{v}}^n\|$ as a function of J for several values of $n\delta t$. On the top panels : $(\alpha, x_0, y_0) = (15, 1, 2)$ so that the derivatives of the solution on the boundary are negligible. On the bottom panels : $(\alpha, x_0, y_0) = (1, 0, 4)$ so that the derivatives of the solution on the boundary are not negligible.

$|1+\delta t\lambda_\ell| \le 1$. Therefore, using (11), we may compute

$$\begin{aligned}
\sum_{\ell=1}^{J-1}\left|\delta t\sum_{k=0}^{n-1}(1+\delta t\lambda_\ell)^k\right|^2 &= \sum_{\ell=1}^{J-1}\left|\delta t\frac{1-(1+\delta t\lambda_\ell)^n}{1-(1+\delta t\lambda_\ell)}\right|^2\\
&\le \sum_{\ell=1}^{J-1}\left|\frac{2\delta t}{-\delta t\lambda_\ell}\right|^2\\
&\le 4\sum_{\ell=1}^{J-1}\frac{1}{\lambda_\ell^2}\\
&\le 4\sum_{\ell=1}^{J-1}\frac{\delta x^4}{16\sin\left(\frac{\ell\pi}{2J}\right)^4}.
\end{aligned}$$

Using the concavity of the sine function, one has classically that for $s\in(0,\pi/2)$, $2s/\pi \le \sin(s)$. This implies

$$\begin{aligned}
\sum_{\ell=1}^{J-1}\left|\delta t\sum_{k=0}^{n-1}(1+\delta t\lambda_\ell)^k\right|^2 &\le \frac{1}{4}\delta x^4\sum_{\ell=1}^{J-1}\frac{\pi^4}{2^4\left(\frac{\ell\pi}{2J}\right)^4}\\
&\le \frac{1}{4}\delta x^4 J^4\sum_{\ell=1}^{J-1}\frac{1}{\ell^4}.
\end{aligned}$$

Since $\delta x J = L\times(J/(J-1)) \le 2L$ and the sum in the right-hand side above is bounded by $\zeta(4)=\pi^4/90$, the proof is complete. $\square$

Remark 6.1. *For the conclusion of Proposition 3.4 to hold, the proof above indicates that it is sufficient that $\sum_{\ell=1}^{J-1}1/\lambda_\ell^2$ can be bounded independently of $J\ge 2$. This hypothesis, in which δt plays no role, is related to a spectral gap condition on P_δ and how well the eigenvalues of P_δ approximate the spectrum of P. Indeed, the nonzero eigenvalues of P are $-(p\pi/L)^2$ for $p\ge 1$ and the series $(1/(-(p\pi/L)^2)^2)_{p\ge 1}$ is also convergent.*

6.2 Two lemmas of numerical quadrature

Lemma 6.2 (Numerical integration over the spectrum of P_δ)**.** *Let $L>0$ be fixed. There exists a constant $C>0$ such that for all $J\ge 2$, $m\in\mathbb{N}\setminus\{0\}$, and all $\alpha>0$,*

$$\delta x\sum_{\ell=1}^{J-1}\mathrm{e}^{-\alpha m\sin^2\left(\frac{\ell\pi}{J}\right)} \le C\frac{1}{\sqrt{m\alpha}}. \tag{60}$$

Proof. Let us first observe that

$$\forall \ell\in\{1,\cdots,J-1\},\qquad \sin\left(\frac{\ell\pi}{J}\right)=\sin\left(\frac{(J-\ell)\pi}{J}\right). \tag{61}$$

Hence, if $J-1$ is even, then

$$\delta x\sum_{\ell=1}^{J-1}\mathrm{e}^{-\alpha m\sin^2\left(\frac{\ell\pi}{J}\right)} = 2\delta x\sum_{\ell=1}^{\frac{J-1}{2}}\mathrm{e}^{-\alpha m\sin^2\left(\frac{\ell\pi}{J}\right)}, \tag{62}$$

and, if $J-1$ is odd, then

$$\delta x\sum_{\ell=1}^{J-1}\mathrm{e}^{-\alpha m\sin^2\left(\frac{\ell\pi}{J}\right)}=2\delta x\sum_{\ell=1}^{\frac{J}{2}-1}\mathrm{e}^{-\alpha m\sin^2\left(\frac{\ell\pi}{J}\right)}+\delta x\mathrm{e}^{-\alpha m\sin^2\left(\frac{\frac{J}{2}\pi}{J}\right)}. \tag{63}$$

Assume that $J-1$ is even. In this case, for $\ell\in\{1,\cdots,(J-1)/2\}$, we have $\ell\pi/J\in(0,\pi/2)$, and hence

$$\frac{2}{\pi}\frac{\ell\pi}{J}\leq\sin\left(\frac{\ell\pi}{J}\right). \tag{64}$$

This implies, using (62),

$$\begin{aligned}
\delta x\sum_{\ell=1}^{J-1}\mathrm{e}^{-\alpha m\sin^2\left(\frac{\ell\pi}{J}\right)} &\leq 2\frac{L}{J-1}\sum_{\ell=1}^{\frac{J-1}{2}}\mathrm{e}^{-\alpha m\left(\frac{2\ell}{J}\right)^2}\\
&\leq L\frac{J}{J-1}\sum_{\ell=1}^{\frac{J-1}{2}}\frac{2}{J}\mathrm{e}^{-\alpha m\left(\frac{2\ell}{J}\right)^2}\\
&\leq L\frac{J}{J-1}\sum_{\ell=1}^{\frac{J-1}{2}}\int_{(\ell-1)/J}^{\ell/J}\mathrm{e}^{-\alpha mx^2}\mathrm{d}x\\
&\leq L\frac{J}{J-1}\int_0^{\frac{1}{2}}\mathrm{e}^{-\alpha mx^2}\mathrm{d}x\\
&\leq \frac{L}{\sqrt{\alpha m}}\frac{J}{J-1}\int_0^{\frac{\sqrt{\alpha m}}{4}}\mathrm{e}^{-u^2}\mathrm{d}u\\
&\leq \frac{C}{\sqrt{\alpha m}},
\end{aligned}$$

with $C=2L\int_0^{+\infty}\mathrm{e}^{-u^2}\mathrm{d}u$.

Assume that $J-1$ is odd. In this case for all $\ell\in\{1,\cdots,J/2\}$, we have $\ell\pi/J\in(0,\pi/2]$. Hence, (64) is valid for such ℓ. Using (63), we infer

$$\begin{aligned}
\delta x\sum_{\ell=1}^{J-1}\mathrm{e}^{-\alpha m\sin^2\left(\frac{\ell\pi}{J}\right)} &\leq 2\frac{L}{J-1}\sum_{\ell=1}^{\frac{J}{2}-1}\mathrm{e}^{-\alpha m\left(\frac{2\ell}{J}\right)^2}+2\frac{L}{J-1}\mathrm{e}^{-\alpha m}\\
&\leq L\frac{J}{J-1}\sum_{\ell=1}^{\frac{J}{2}}\frac{2}{J}\mathrm{e}^{-\alpha m\left(\frac{2\ell}{J}\right)^2}\\
&\leq L\frac{J}{J-1}\sum_{\ell=1}^{\frac{J}{2}}\int_{(\ell-1)/J}^{\ell/J}\mathrm{e}^{-\alpha mx^2}\mathrm{d}x\\
&\leq L\frac{J}{J-1}\int_0^{\frac{1}{2}}\mathrm{e}^{-\alpha mx^2}\mathrm{d}x\\
&\leq \frac{L}{\sqrt{\alpha m}}\frac{J}{J-1}\int_0^{\frac{\sqrt{\alpha m}}{4}}\mathrm{e}^{-u^2}\mathrm{d}u\\
&\leq \frac{C}{\sqrt{\alpha m}},
\end{aligned}$$

with the same C as above. □

Remark 6.3. *Note that the bound (60) allows to carry on the computations of order since it provides (for $\alpha = \delta t/\delta x^2$) a uniform bound on the sum in the left-hand side that*

- *tends to 0 as m tends to $+\infty$*
- *does not depend on δt except via the CFL number α.*

In some sense, it is uniform in $\delta t, \delta x$ in the CFL region (and still tends to 0 as m tends do $+\infty$). In contrast, using just a spectral gap in the eigenvalues of P_δ would lead to a bound of the form

$$\begin{aligned}\delta x \sum_{\ell=1}^{J-1} |1+\delta t \lambda_\ell|^m &\leq \delta x \sum_{\ell=1}^{J-1} \left(1 - \frac{\pi^2}{L^2}\delta t\right)^m \\ &\leq L\left(1-\frac{\pi^2}{L^2}\delta t\right)^m \\ &\leq L \mathrm{e}^{-\frac{\pi^2}{L^2} m \delta t},\end{aligned}$$

which, when we take the supremum in $\delta t, \delta x$ in the CFL region yields $L \times 1$ and no longer tends to 0 when m tends to ∞.

Lemma 6.4. *Let $L > 0$ be fixed. There exists $C > 0$ such that for all $\delta t \in (0,1)$, $p \geq 1$, $n \geq 1$,*

$$\delta t^2 \sum_{\substack{(k_1,k_2)\in\{0,\cdots,n-1\}^2 \\ 2n-2-k_1-k_2 \geq 1}} \mathrm{e}^{-p^2 \frac{\pi^2}{L^2}(k_1+k_2)\delta t} \frac{1}{\sqrt{(2n-2-k_1-k_2)\delta t}} \leq C.$$

Proof. Because of the monotonicity and positivity of the terms in the sum, it is sufficient to prove the result for $p = 1$. Hence, we assume $p = 1$. The sum is empty if $n = 1$, so any positive C will work. We assume $n \geq 2$. Let us denote by

$$\mathcal{E}_n = \{(k_1,k_2) \in \{0,\cdots,n-1\}^2 \quad | \quad 0 \leq k_1 + k_2 \leq 2n-3\},$$

which we split into

$$\mathcal{E}_n^1(\delta t) = \{(k_1,k_2) \in \mathcal{E} \quad | \quad (2n-2-k_1-k_2)\delta t \geq 1\},$$

and

$$\mathcal{E}_n^2(\delta t) = \{(k_1,k_2) \in \mathcal{E} \quad | \quad (2n-2-k_1-k_2)\delta t < 1\}.$$

Setting

$$\mathcal{T}_n(\delta t) = \{(s,t) \in \mathbb{R}^2 \quad | \quad s > 0,\ t > 0,\ s + t < 2(n-1)\delta t\},$$

and $f_{(n-1)\delta t}(s,t) = \mathrm{e}^{-\frac{\pi^2}{L^2}(s+t)} \frac{1}{\sqrt{2(n-1)\delta t - (s+t)}}$, we can carry estimates as follows:

$$\begin{aligned}&\delta t^2 \sum_{(k_1,k_2)\in\mathcal{E}_n} f_{(n-1)\delta t}(k_1\delta t, k_2\delta t) \\ &\leq \delta t^2 \sum_{(k_1,k_2)\in\mathcal{E}_n^1(\delta t)} f_{(n-1)\delta t}(k_1\delta t, k_2\delta t) + \delta t^2 \sum_{(k_1,k_2)\in\mathcal{E}_n^2(\delta t)} f_{(n-1)\delta t}(k_1\delta t, k_2\delta t) \\ &\leq \delta t^2 \sum_{(k_1,k_2)\in\mathcal{E}_n^1(\delta t)} \mathrm{e}^{-\frac{\pi^2}{L^2}(k_1+k_2)\delta t} + \delta t^2 \sum_{(k_1,k_2)\in\mathcal{E}_n^2(\delta t)} f_{(n-1)\delta t}(k_1\delta t, k_2\delta t). \qquad (65)\end{aligned}$$

For the first term in the right-hand side of (65), we have

$$\begin{aligned}
\delta t^2 \sum_{(k_1,k_2)\in\mathcal{E}_n^1(\delta t)} \mathrm{e}^{-\frac{\pi^2}{L^2}(k_1+k_2)\delta t} &\leq \delta t^2 \sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} \mathrm{e}^{-\frac{\pi^2}{L^2}(k_1+k_2)\delta t} \\
&\leq \left(\delta t \sum_{k_1=0}^{n-1} \mathrm{e}^{-\frac{\pi^2}{L^2}k_1\delta t}\right)\left(\delta t \sum_{k_2=0}^{n-1} \mathrm{e}^{-\frac{\pi^2}{L^2}k_2\delta t}\right) \\
&\leq \left(\frac{\delta t}{1-\mathrm{e}^{-\frac{\pi^2}{L^2}\delta t}}\right)^2 .
\end{aligned}$$

This last term does not depend on $n \geq 2$ and is bounded independently of $\delta t \in (0,1)$. For the second term in (65), observe that, for $(k_1,k_2) \in \mathcal{E}_n^2(\delta t)$, we have $(2n-2)\delta t - 1 < (k_1+k_2)\delta t$, and hence

$$0 \leq f_{(n-1)\delta t}(k_1\delta t, k_2\delta t) \leq \mathrm{e}^{-\frac{\pi^2}{L^2}((2n-2)\delta t-1)} \frac{1}{\sqrt{(2n-2)\delta t - k_1\delta t - k_2\delta t}}.$$

Since we also have, using the monotonicity of the function $t \mapsto 1/\sqrt{(2n-2)\delta t - t}$ over $(-\infty,(2n-2)\delta t)$,

$$\frac{\delta t^2}{2}\frac{1}{\sqrt{(2n-2)\delta t - k_1\delta t - k_2\delta t}} \leq \int_{(k_1\delta t,k_2\delta t)+\mathcal{T}_{3/2}(\delta t)} \frac{1}{\sqrt{(2n-2)\delta t - s - t}}\mathrm{d}t\mathrm{d}s,$$

we may write

$$\begin{aligned}
&\delta t^2 \sum_{(k_1,k_2)\in\mathcal{E}_n^2(\delta t)} f_{(n-1)\delta t}(k_1\delta t, k_2\delta t) \\
&\leq 2\,\mathrm{e}^{\frac{\pi^2}{L^2}-\frac{\pi^2}{L^2}(2n-2)\delta t} \sum_{(k_1,k_2)\in\mathcal{E}_n^2(\delta t)} \int_{(k_1\delta t,k_2\delta t)+\mathcal{T}_{3/2}(\delta t)} \frac{1}{\sqrt{(2n-2)\delta t - s - t}}\mathrm{d}t\mathrm{d}s \\
&\leq 2\,\mathrm{e}^{\frac{\pi^2}{L^2}-\frac{\pi^2}{L^2}(2n-2)\delta t} \sum_{(k_1,k_2)\in\mathcal{E}} \int_{(k_1\delta t,k_2\delta t)+\mathcal{T}_{3/2}(\delta t)} \frac{1}{\sqrt{(2n-2)\delta t - s - t}}\mathrm{d}t\mathrm{d}s \\
&\leq 2\,\mathrm{e}^{\frac{\pi^2}{L^2}-\frac{\pi^2}{L^2}(2n-2)\delta t} \int_{\mathcal{T}_n(\delta t)} \frac{1}{\sqrt{(2n-2)\delta t - (s+t)}}\mathrm{d}t\mathrm{d}s \\
&\leq 2\,\mathrm{e}^{\frac{\pi^2}{L^2}-\frac{\pi^2}{L^2}(2n-2)\delta t} \int_0^{2(n-1)\delta t}\int_{\mathsf{v}}^{2(n-1)\delta t} \frac{1}{\sqrt{(2n-2)\delta t - u}}\mathrm{d}u\mathrm{d}v \\
&\leq 2\,\mathrm{e}^{\frac{\pi^2}{L^2}-\frac{\pi^2}{L^2}(2n-2)\delta t} \int_0^{2(n-1)\delta t} 2\sqrt{(2n-2)\delta t - v}\mathrm{d}v \\
&\leq 4\,\mathrm{e}^{\frac{\pi^2}{L^2}-\frac{\pi^2}{L^2}(2n-2)\delta t}\frac{2}{3}\left((2n-2)\delta t\right)^{3/2} \\
&\leq \frac{8}{3}\mathrm{e}^{\frac{\pi^2}{L^2}-\frac{\pi^2}{L^2}(2n-2)\delta t}\left((2n-2)\delta t\right)^{3/2}.
\end{aligned}$$

This last term is bounded independently of $n \geq 2$ and $\delta t > 0$ by $C = (8\mathrm{e}^{\frac{\pi^2}{L^2}}/3) \times \sup_{x\in(0,+\infty)} x^{3/2}\mathrm{e}^{-\frac{\pi^2}{L^2}x}$. This concludes the proof of the lemma. $\square$

Lemma 6.5. *Let $L > 0$ be fixed. There exists a constant $C > 0$ such that for all $J \geq 2$ and all $v \in H^1(0,L)$,*

$$\|\Pi_{\delta x} v\|_{\ell^2}^2 \leq C\|v\|_{H^1}^2.$$

Proof. Let $v \in H^1(0,L)$ be fixed. For $j \in \{0,\cdots,J-2\}$, and $x \in [x_j, x_{j+1}]$,

$$v^2(x) = v^2(x_j) + \int_{x_j}^{x} \partial_x v^2(s)\mathrm{d}s. \tag{66}$$

Integrating over $[x_j, x_{j+1}]$, we obtain

$$\int_{x_j}^{x_{j+1}} v^2(x)\mathrm{d}x = \delta x\, v^2(x_j) + 2\int_{x_j}^{x_{j+1}} \int_{x_j}^{x} v(s)\partial_x v(s)\mathrm{d}s\mathrm{d}x.$$

In particular, for all $j \in \{0,\cdots,J-2\}$,

$$\begin{aligned}\left|\int_{x_j}^{x_{j+1}} v^2(x)\mathrm{d}x - \delta x\, v^2(x_j)\right| &\leq \int_{x_j}^{x_{j+1}} \int_{x_j}^{x_{j+1}} \left(v^2(s) + (\partial_x v)^2(s)\right)\mathrm{d}s\mathrm{d}x \\ &\leq \delta x \int_{x_j}^{x_{j+1}} \left(v^2(s) + (\partial_x v)^2(s)\right)\mathrm{d}s. \end{aligned}\tag{67}$$

Moreover, we have, for $x \in [x_{J-2}, x_{J-1}]$,

$$v^2(x) = v^2(x_{J-1}) + 2\int_{x_{J-1}}^{x} v(x)\partial_x v(s)\mathrm{d}s.$$

Integrating over $[x_{J-2}, x_{J-1}]$, we obtain

$$\int_{x_{J-2}}^{x_{J-1}} v^2(x)\mathrm{d}x = \delta x\, v^2(x_{J-1}) + 2\int_{x_{J-2}}^{x_{J-1}} \int_{x_{J-1}}^{x} v(x)\partial_x v(s)\mathrm{d}s\mathrm{d}x.$$

This yields

$$\left|\int_{x_{J-2}}^{x_{J-1}} v^2(x)\mathrm{d}x - \delta x\, v^2(x_{J-1})\right| \leq \delta x \int_{x_{J-2}}^{x_{J-1}} \left(v^2(s) + (\partial_x v)^2(s)\right)\mathrm{d}s. \tag{68}$$

Summing (67) with respect to j in $\{0,\cdots,J-2\}$ and (68), and then dividing by L, we obtain by triangle inequality

$$\left|\frac{1}{L}\int_0^L v^2(x)\mathrm{d}x + \frac{1}{L}\int_{x_{J-2}}^{x_{J-1}} v^2(x)\mathrm{d}x - \frac{1}{J-1}\sum_{j=0}^{J-1} v^2(x_j)\right| \leq \frac{2}{L}\delta x \int_0^L \left(v^2(x) + (\partial_x v)^2(x)\right)\mathrm{d}x.$$

This implies, by inverse triangle inequality,

$$\begin{aligned}\|\Pi_{\delta x} v\|_{\ell^2}^2 &= \frac{1}{J}\sum_{j=0}^{J-1} v^2(x_j) \\ &\leq \frac{J-1}{J}\frac{1}{J-1}\sum_{j=0}^{J-1} v^2(x_j) \\ &\leq \frac{J-1}{J}\, 2\,(1+\delta x)\,\|v\|_{H^1}^2.\end{aligned}$$

Since $(J-1)/J \leq 1$ and $\delta x = L/(J-1)$ is bounded independently of $J \geq 2$, this proves the lemma. $\square$

Remark 6.6. *In the context of the discretization of the homogeneous Fokker–Planck equation*

$$\partial_t u = -(-\partial_v + v)\partial_v u, \tag{69}$$

with homogeneous Neumann boundary conditions over a finite interval $(0, L)$*, one obtains a discrete (in velocity) problem of the form*

$$\partial_t u = \mathsf{P}_\delta u, \tag{70}$$

where P_δ *typically is a nonpositive symmetric square matrix of size* $J \geq 2$*. Provided one can show that one can number the eigenvalues* $(\lambda_\ell)_{0\leq\ell\leq J-1}$ *of* P_δ *in such a way that one has*

$$\forall \delta t, \delta v > 0, \quad \forall \ell \in \{0, \cdots, J-1\}, \qquad |1 + \delta t \lambda_\ell| \leq \mathrm{e}^{-\frac{\delta t}{\delta v^2} g(\ell/J)}, \tag{71}$$

for some nonnegative continuous function g *over* $[0, 1]$ *(that may now depend on* δv*) such that* $g(0) = 0$ *and*

$$\forall \delta v > 0, \quad \forall j \in \{1, \cdots, \lfloor J/2 \rfloor\}, \qquad c\left(\frac{\ell}{J}\right)^2 \leq g\left(\frac{\ell}{J}\right) \leq g\left(\frac{J-\ell}{J}\right), \tag{72}$$

for some $c > 0$ *(that does not depend on* δv*), one gets an analogue of Theorem 2.12 for the explicit Euler method applied to the discretized (in velocity) Fokker–Planck equation* (70) *when compared to the projection of the exact solution of* (69) *: the order of that method is uniform in time. Indeed, the error analysis is the same for all the terms and follows the same lines. In particular, for the terms in* $\mathcal{L}^1$*, the hypothesis* (71) *plays the role of Proposition 3.2 with* $g(x) = \sin^2(\pi x)$*, and* (71) *ensures that an analogue of* (60) *makes a similar result to Lemma 6.2 true.*

6.3 Comparison with an existing longtime numerical analysis framework

This section is devoted to explaining the reasons why the longtime analysis of the numerical scheme (12) applied to the linear heat equation with Neumann boundary conditions (1) does not fit usual longtime numerical analysis frameworks. We take for example the framework developed in [20]. For readability, we use the convention, in this section, that in all the equalities and inequalities, left-hand sides use the notations of [20] and right-hand sides correspond to the notations of our paper.

Looking at (2.1) in [20], we have $f \equiv 0$, $g \equiv 0$, $A = P$. Moreover, we have $\tau = \delta t$ and $h = \delta x$. Using the notation of (2.2), we have $B_{h,\tau} = \mathsf{Id}$, $C_{h,\tau} = \mathsf{Id} + \delta t \mathsf{P}_\delta$, $g^n_{h,\tau} = 0$. In (2.4), we have $L_{h,\tau}(p_h(u(n\tau))) = (\varepsilon^n_1 + \varepsilon^n_2)/\delta t$ (with the notation introduced just before our error expansion formula (21)). With standard regularity assumptions implied by our Hypothesis (6), we have that $\|\varepsilon^n_2/\delta t\|_{\ell^2}$ is a $\mathcal{O}(\delta t)$ where the constant in the $\mathcal{O}$ does not depend on n, δt and δx under the CFL condition (13). However, the lack of consistency described in Section 2.4 shows that $\|\mathcal{L}^1_\delta u(n\delta t)\|_{\ell^2}$ behaves as $\mathcal{O}(\delta x^{1/2})$ (only 2 nonzero terms of order 1 at the boundary) and $\|\delta x^2 \mathcal{L}^2_\delta u(n\delta t)\|_{\ell^2}$ is of order $\mathcal{O}(\delta x^2)$, where the constants do not depend either on n, δt and δx under the CFL condition (13). Moreover, for a general solution, these orders cannot be improved in general. This implies that $\|\varepsilon^n_1/\delta t\|_{\ell^2}$ behaves as $\mathcal{O}(\delta x^{1/2})$. Using (2.6) in [20], and the computations above, we infer that $S_{h,\tau} \geq C\delta x^{1/2}$ for some positive C. In particular, applying Theorem 2.1 of [20] yields a uniform bound on $e^n_{h,\tau}$ that is, at best, $\mathcal{O}(\delta x^{1/2})$ (see relation (2.12) in [20]). This is much coarser than our result which provides a uniform bound of size $\mathcal{O}(\delta x)$ (see (20) in Theorem 2.12).

Ethical statement: G.D. is supported by the Inria project-team PARADYSE and the Labex CEMPI (ANR-11-LABX-0007-01). Conflict of Interest: The authors declare that they have no conflict of interest.

References

[1] Pavel B. Bochev and Richard B. Lehoucq. On the finite element solution of the pure Neumann problem. SIAM Rev., 47:50–66, 2005.

[2] Claire Chainais-Hillairet, Maxime Herda, Simon Lemaire, and Julien Moatti. Long-time behaviour of hybrid finite volume schemes for advection–diffusion equations: linear and nonlinear approaches. Numerische Mathematik, 151:963–1016, 2022.

[3] Chuan-Miao Chen, Stig Larsson, and Nai-Ying Zhang. Error Estimates of Optimal Order for Finite Element Methods with Interpolated Coefficients for the Nonlinear Heat Equation. IMA Journal of Numerical Analysis, 9(4):507–524, 10 1989.

[4] Jérôme Droniou, Robert Eymard, Thierry Gallouët, Cindy Guichard, and Raphaèle Herbin. The Gradient Discretisation Method. Mathématiques et Applications. Springer Cham, 2018.

[5] Jérôme Droniou, Neela Nataraj, and Devika Shylaja. Numerical analysis for the pure Neumann control problem using the gradient discretisation method. Computational Methods in Applied Mathematics, 18(4):609–637, 2018.

[6] Guillaume Dujardin, Frédéric Hérau, and Pauline Lafitte. Coercivity, hypocoercivity, exponential time decay and simulations for discrete Fokker–Planck equations. Numerische Mathematik, 144(3):615–647, 2020.

[7] Ronald D. Haynes and Khaled Mohammad. Fully discrete schwarz waveform relaxation analysis for the heat equation on a finite spatial domain. ESAIM: M2AN, 57(4):2397–2426, 2003.

[8] Hussain A. Ibdah, Cecilia F. Mondaini, and Edriss S. Titi. Fully discrete numerical schemes of a data assimilation algorithm: uniform-in-time error estimates. IMA Journal of Numerical Analysis, 40(4):2584–2625, 11 2019.

[9] Mélanie Inglard, Frédéric Lagoutière, and Hans Henrik Rugh. Ghost solutions with centered schemes for one-dimensional transport equations with Neumann boundary conditions. Annales de la Faculté des sciences de Toulouse : Mathématiques, Ser. 6, 29(4):927–950, 2020.

[10] Bangti Jin, Raytcho Lazarov, and Zhi Zhou. Numerical methods for time-fractional evolution equations with nonsmooth data: A concise overview. Computer Methods in Applied Mechanics and Engineering, 346:332–358, 2019.

[11] Heinz-Otto Kreiss, N. Anders Petersson, and Jacob Yström. Difference approximations of the Neumann problem for the second order wave equation. SIAM Journal on Numerical Analysis, 42(3):1292–1323, 2004.

[12] Randall J. LeVeque. Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems. SIAM, 2007.

[13] Sylvain Maire and Etienne Tanré. Monte Carlo approximations of the Neumann problem. Monte Carlo Methods and Applications, 19(3):201–236, 2013.

[14] Cecilia F. Mondaini and Edriss S. Titi. Uniform-in-time error estimates for the postprocessing galerkin method applied to a data assimilation algorithm. SIAM Journal on Numerical Analysis, 56(1):78–110, 2018.

[15] Lionel Sainsaulieu. Calcul scientifique: cours et exercices corrigés pour le 2e cycle et les écoles d'ingénieurs. Masson, 1996.

[16] J. M. Sanz-Serna and A. M. Stuart. A note on uniform in time error estimates for approximations to reaction-diffusion equations. IMA Journal of Numerical Analysis, 12(3):457–462, 07 1992.

[17] Suraj Shankar. 2D Laplace equation. `https://fr.mathworks.com/matlabcentral/fileexchange/38091-2d` Updated 10 Sep 2012.

[18] Jie Shen. Long time stability and convergence for fully discrete nonlinear galerkin methods. Applicable Analysis, 38:201–229, 01 1990.

[19] James William Thomas. Numerical partial differential equations: finite difference methods, volume 22. Springer Science & Business Media, 1995.

[20] Hai-jun Wu and Rong-hua Li. The stability and convergence of computing long-time behaviour. Journal of Computational Mathematics, pages 397–418, 1999.