

ALIASING AND TWO-DIMENSIONAL WELL-BALANCED FOR DRIFT-DIFFUSION EQUATIONS ON SQUARE GRIDS

LAURENT GOSSE

ABSTRACT. A notion of “2D well-balanced” for drift-diffusion is proposed. Exactness at steady-state, typical in 1D, is weakened by aliasing errors when deriving “truly 2D” numerical fluxes from local Green’s function. A main ingredient for proving that such a property holds is the optimality of the trapezoidal rule for periodic functions. In accordance with practical evidence, a “Bessel scheme” previously introduced in [SIAM J. Numer. Anal. 56 (2018), pp. 2845–2870] is shown to be “2D well-balanced” (along with former algorithms known as “discrete weighted means” or “tailored schemes”). Some L^2 stability estimates are established, too.

1. INTRODUCTION AND CONTEXTUALIZATION

Following [10], we hereafter pursue the study of “genuinely two-dimensional” numerical approximations of drift-diffusion equations, of the form

$$(1) \quad \partial_t u(t, x, y) - \nabla \cdot (\varepsilon \nabla u - u \mathbf{V}(x, y)) = 0 \quad \text{in } \Omega \subset \mathbb{R}^2, \quad \varepsilon > 0,$$

where Ω is, most often, the square domain $(0, 1)^2$ and $\partial\Omega$ its boundary. Convenient boundary conditions supplement (1), like, e.g., Dirichlet or Neumann. A linear model like (1) may be useful for solving incompressible Navier-Stokes equations,

$$(2) \quad \partial_t \omega(t, x, y) - \nabla \cdot \left(\frac{\nabla \omega}{Re} - \omega \nabla^\perp \psi \right) = 0, \quad -\Delta \psi = \omega, \quad 0 < Re = \frac{1}{\varepsilon},$$

as soon as a “splitting algorithm”, where the evolution of ω is separated from the Poisson equation for ψ is chosen [16]. The stream function ψ is related to vorticity $\omega = \nabla^\perp \cdot \vec{U}$, where $\vec{U} = (u, v)$ stands for the (divergence-free) fluid’s velocity.

1.1. A puzzling observation about 2D grid-sensitivity. One motivation for the derivation of the 2D scheme presented in [10] was somehow “trying to extend well-balanced numerical methods, nowadays standard for 1D hyperbolic systems of balance laws, toward 2D convection-diffusion problems”. In particular, there was a real question about the possibility of having, in 2D, error bounds as good as the ones proved in [2] for 1D problems. Accordingly, the scheme in [10] being able to handle the 2D incompressible Navier-Stokes system (2), we set it up on the standard benchmark of the “lid-driven cavity”, for which exact values are available with various Reynolds numbers in, e.g., [16]. To assess accuracy at steady-state, for which we already have at hand the values given in [16], we compared our results

Received by the editor March 8, 2018, and, in revised form, November 18, 2018.

2010 *Mathematics Subject Classification.* Primary 35K15, 65M12, 76D05, 76R50.

The support of Italian Minister of Instruction, University and Research (MIUR) through PRIN Project 2017, entitled “Innovative numerical methods for evolutionary partial differential equations and applications” #2017KKJP4X is acknowledged.

with the ones generated by the classic scheme of Arakawa [5], which is known for dissipating enstrophy and being “mimetic” in the sense of [26]. For $Re = 7500$ in (2), we observed the results displayed in Figure 1. Clearly, they were puzzling

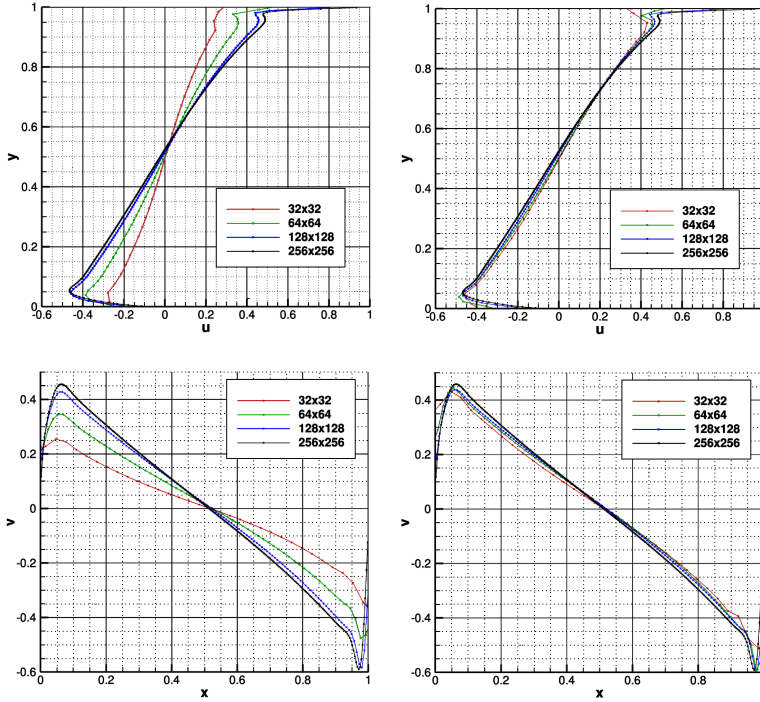


FIGURE 1. Sensitivity of Arakawa (left) and Bessel (right) schemes to the grid size $1/\Delta x$ for the lid-driven cavity at Reynolds=7500.

because the steady-state curves for both horizontal and vertical velocities appeared to be “nearly independent” of the grid parameter, $\Delta x > 0$, oppositely to Arakawa’s scheme. Such a behavior is strongly reminiscent of the so-called “well-balanced” property (a term coined in [21]; see [17, Chap. 4] for historical perspectives) in one space dimension; roughly speaking, it consists of building the numerical scheme in such a way it can exactly preserve the stationary regimes of the underlying problem whatever the grid-size (another way is to say that these schemes have no artificial viscosity at steady-state). However, it is still a challenge to mathematically define an equivalent of such a property in two (or more) spatial dimensions [1], left apart proving it for the “2D Bessel scheme” of [10], when applied to (2). In this paper, we restrict ourselves to linear drift-diffusion equations (1) and pursue two goals:

- To define rigorously a notion of “2D well-balanced” property for this class of PDE’s such that it’s possible to prove that the scheme of [10] satisfies it;
- To prove that this (semi-discrete) scheme is endowed with L^2 bounds, in a similar way to the mimetic one originally proposed by Arakawa in [5].

In one space dimension, well-balanced numerical fluxes usually follow from a (local) integration of stationary equations, i.e., a boundary-value problem (BVP); see, e.g., [2, 17, 20]. In more dimensions, in order to retrieve any numerical flux from a BVP,

one must reconstruct “co-dimension one boundary data” out of a limited amount of grid samples. Such a sampling step (see, e.g., [27]) is clearly unnecessary in 1D. Accordingly, **a scheme is “2D well-balanced” as soon as sampling noise dominates its local truncation error when initial data is at steady-state.** The “well-balanced kinetic schemes” (see [17, Ch. 9-10]) rely on an S -matrix which proceeds by interpolating discrete data on so-called “Case elementary solutions”.

Definition 1. On a uniform Cartesian grid characterized by $\Delta x = \Delta y$, let $u_{i,j}(t)$ stand for a numerical approximation of $u(t, i\Delta x, j\Delta y)$, solution of (1) associated to initial data $u^0 \in H^s(\Omega)$, $\Omega \subset \mathbb{R}^2$ Lipschitz and convex. The numerical scheme is “2D well-balanced” as soon as its (initial) Local Truncation Error (LTE) satisfies

$$(3) \quad \forall (i, j) \in \mathbb{Z}^2, \quad s > \frac{3}{2}, \quad |u_{i,j}(\Delta t) - u^0(x_i, y_j)| \leq O\left(\Delta x^{\min(3, s - \frac{1}{2}) - \chi(\Delta x \geq \varepsilon)}\right),$$

when u^0 is a stationary state for (1), that is, $\nabla \cdot (\varepsilon \nabla u^0 - \mathbf{V}(x, y)u^0) = 0$.

A few comments about (3) and regularity assumptions are as follows:

- A factor $\frac{1}{2}$ comes from 2D functions in H^s having circular traces in $H^{s-\frac{1}{2}}$;
- If $\Delta x \geq \varepsilon$, CFL restrictions pass from parabolic $\Delta t = O(\Delta x^2)$ to hyperbolic;
- Condition $s > \frac{3}{2}$ matches the typical H^2 regularity in square domains;
- Finally, $u \in H^{\frac{7}{2}}(\Omega)$ is substantially weaker than what is assumed when deriving stability estimates from Taylor expansions. Indeed, in, e.g., [3, 4, 37], exact solutions are asked to (roughly) belong to C^5 . Presently, Sobolev embeddings in 2D ensure that, for any Lipschitz domain Ω , $H^{\frac{7}{2}} \subset C^{2, \frac{1}{2}}$.

Estimate (3) expresses that a general $C^{2, \frac{1}{2}}$ steady-state for the linear problem (1) is quite stable. Yet, coming back to the numerical example in Figure 1, such a result cannot be directly applied because Navier-Stokes equations are nonlinear; however, we believe that (3) still induces, within a splitting strategy (the evolution of ω is dissociated from the Poisson equation), a better stability of numerical steady-states.

1.2. Main results and outline of the paper. Our main result is as follows.

Theorem 1. *Given $u(t, \cdot, \cdot) \in H^s(\mathbb{R}^2)$, $s \geq 2$, a solution of (1) for a smooth and bounded velocity field $\mathbf{V}(x, y)$, the “Bessel scheme” written in [10, §3-4]*

- (1) *is “2D well-balanced” in the sense of Definition 1;*
- (2) *it preserves the total mass:*

$$\forall t > 0, \quad \frac{d}{dt} \left(\sum_{(i,j) \in \mathbb{Z}^2} u_{i,j}(t) \right) = 0;$$

- (3) *if $\mathbf{V} \in \mathbb{R}^2$ is a constant vector, then*

$$\forall \varepsilon > 0, \quad \frac{d}{dt} \left(\sum_{i,j} |u_{i,j}(t)|^2 \right) \leq 0;$$

- (4) *instead, if $\mathbf{V}(x, y)$ is only divergence-free, then*

$$\forall \varepsilon > 0, \quad \frac{d}{dt} \left(\Delta x^2 \sum_{i,j} |u_{i,j}(t)|^2 \right) \leq 2\Delta x^2 \|\mathbf{V}\|_\infty \|u(t)\|_{H^2(\mathbb{R}^2)}^2.$$

At this level, these L^2 estimates imply that,

- If $\mathbf{V} \in \mathbb{R}^2$ is constant, the 2D Bessel scheme dissipates the L^2 norm like the Arakawa scheme does, so that the Bessel scheme is essentially mimetic. Meanwhile, it appears to capture steady-states in a better way (see Figure 1), in the sense that it displays less sensitivity to the grid-parameter $\Delta x > 0$.
- When \mathbf{V} varies in space, but is divergence-free, and the exact solution is smooth enough, the Bessel 2D scheme can be considered as reliable for times $t \leq O(\Delta x^{-2})$, uniformly in $\varepsilon > 0$.

The remainder of the paper is essentially devoted to proving Theorem 1: Assertion (1) is studied in Section 2. A crucial ingredient is Lemma 2, proved in Appendix A, in the spirit of [12, 31, 38, 40]. Proposition 2 states that for initial data which are steady-states for (1) belonging to H^s , $s > \frac{3}{2}$, the requirements of Definition 1 are met. A closely related property for the so-called “discrete weighted mean” methods (see [3, 13, 37]) and the “tailored schemes” [22] is studied in Appendix B. Section 3 contains the proofs of Assertions (2), (3), and (4). After having recalled the computations in 1D for Scharfetter-Gummel’s scheme [34], the case of constant velocities is studied in §3.3, whereas the more general one of divergence-free velocities is to be found in §3.4. Concluding remarks are given in Section 4.

2. 2D WELL-BALANCED WITHOUT DIMENSIONAL-SPLITTING

This section contains a proof of Assertion (1) in the Main Theorem 1.

2.1. Regularity theory for the stationary problem. When (1) is posed in a bounded domain Ω , it is customary to prescribe conditions on its boundary, the simplest being $u = g$ on $\partial\Omega$. In order to follow results established in [8] for singularly perturbed convection-diffusion equations, it is necessary to recast (1) in a more convenient form. To proceed, let $G(x, y)$ be a smooth function such that $G = g$ on $\partial\Omega$; for instance, it may be its “harmonic extension”, i.e., $\Delta G = 0$ in Ω .

By linearity, let $v = u - G$, so that $\partial_t u = \partial_t v$ and

$$\begin{aligned} -\nabla \cdot (\varepsilon \nabla u - u \mathbf{V}) &= -\nabla \cdot (\varepsilon \nabla (v + G) - (v + G) \mathbf{V}) \\ &= -\varepsilon \Delta v + \mathbf{V} \cdot \nabla v + (\nabla \cdot \mathbf{V})v - \nabla \cdot (\varepsilon \nabla G - G \mathbf{V}), \end{aligned}$$

hence (1) rewrites as,

$$(4) \quad \partial_t v - \varepsilon \Delta v + \mathbf{V} \cdot \nabla v + (\nabla \cdot \mathbf{V})v = f, \quad f = \nabla \cdot (\varepsilon \nabla G - G \mathbf{V}), \quad v = 0 \text{ on } \partial\Omega.$$

The external source simplifies into $f = -\nabla \cdot (G \mathbf{V})$ when G is the harmonic extension of g in Ω . Lemmas 2.1 and 2.2 in [8] can be applied to (4) and yield the following.

Lemma 1 (see [8]). *Assume $\nabla \cdot \mathbf{V} \geq d_0 > 0$, and $f \in H^{k-2}(\Omega)$ with $\partial\Omega$ smooth. Then $v \in H^k \cap H_0^1(\Omega)$ and*

$$\varepsilon^{k-\frac{1}{2}} \|v\|_{H^k} \leq C \|f\|_{H^{k-2}}, \quad (\text{“shift property”}).$$

However, if Ω is only a convex polygon, then $v \in H^2 \cap H_0^1(\Omega)$ and

$$(5) \quad \varepsilon^{\frac{3}{2}} \|v\|_{H^2} + \varepsilon^{\frac{1}{2}} \|v\|_{H^1} + \|v\|_{L^2} \leq C \|f\|_{L^2}.$$

For practical numerical simulations, cases where Ω being a square, hence a convex polygon, are among the most frequent ones; accordingly, v only satisfies (5). More regularity can be hoped for when so-called “compatibility conditions” are met at the edges of the polygon; see [23] and [32, page 236]. The coercivity condition can be met in general thanks to an exponential change of variables; see [32, page 376].

2.2. Gauss-Ortogradski Theorem in disks. Fix $\Delta x = \Delta y > 0$ and consider a uniform Cartesian grid where $x_i = i\Delta x$, $y_j = j\Delta x$ for any couple $(i, j) \in \mathbb{Z}^2$. Given a function $u(t, x, y)$ and $R := \Delta x/\sqrt{2}$, we define

$$\forall (i, j) \in \mathbb{Z}^2, \quad u_{i,j}(t) := \frac{1}{|D_R(i, j)|} \int_{D_R(i, j)} u(t, x, y) \, dx \, dy, \quad t \geq 0,$$

where the disk $D_R(i, j)$ is centered at (x_i, y_j) :

$$D_R(i, j) = \{(x, y) \in \mathbb{R}^2 \text{ such that } |x - x_i|^2 + |y - y_j|^2 \leq R^2\}.$$

If $u(t, x, y)$ solves the continuity equation, then

$$\frac{d u_{i,j}}{dt} = \frac{1}{\pi R^2} \int_{D_R(i, j)} \nabla \cdot (\varepsilon \nabla u - \mathbf{V}(x, y)u) \, dx \, dy.$$

Yet, the Divergence (Gauss-Ortogradski) Theorem yields

$$\frac{d u_{i,j}}{dt} = \frac{1}{\pi R^2} \int_{C_R(i, j)} (\varepsilon \nabla u - \mathbf{V}(x, y)u) \, d\vec{n},$$

where $C_R(i, j) = \partial D_R(i, j)$ is the circle centered in (x_i, y_j) and radius R . Such an integral is easier when passing to polar coordinates, so that

$$(6) \quad \frac{d u_{i,j}}{dt} = \frac{1}{\pi R^2} \int_0^{2\pi} \left(\varepsilon \frac{\partial u}{\partial r}(R, \theta) - \|\mathbf{V}\| \cos(\mu - \theta)u(R, \theta) \right) R \, d\theta,$$

with $\mu \in (0, 2\pi)$ the angle made by the vector \mathbf{V} with respect to the origin of the angular variable θ which parametrizes $C_R(i, j)$. If we had at hand reliable values of both $\frac{\partial u}{\partial r}(R, \theta)$ and $u(R, \theta)$ for $\theta_k = k\pi/2$, we could perform an elementary “method of rectangles” (or “nodal”; see [24]) approximation of the previous integral:

$$(7) \quad \frac{d u_{i,j}}{dt} \simeq \frac{1}{\pi R^2} \frac{\pi R}{2} \sum_{k=0}^3 \left(\varepsilon \frac{\partial u}{\partial r} - \|\mathbf{V}\| \cos(\mu - \theta)u \right) (R, k\pi/2).$$

Such a formula clearly involves “exact fluxes”, denoted $\mathcal{J}_{i,j}$, because it rewrites:

$$(8) \quad \frac{d u_{i,j}}{dt} \simeq \frac{1}{2R} \sum_{k=0}^3 \mathcal{J}_{i,j}(k\pi/2),$$

which is the expression of a conservative (total mass preserving) numerical scheme. Yet, the aforementioned “method of rectangles”, despite its simplicity, is extremely accurate (even optimal) when applied to periodic functions; see, e.g., [28, 31, 38, 40].

Definition 2. Let a function f be T -periodic; then, for $s \geq 0$, $f \in H^s(0, T)$ if

$$\|f\|_{H^s(0, T)}^2 = \sum_{k \in \mathbb{Z}} \left[1 + \left| \frac{2\pi}{T} k \right|^2 \right]^s \hat{f}(k)^2 < \infty, \quad \hat{f}(k) = \int_0^T f(x) \exp(-ik \frac{2\pi x}{T}) \frac{dx}{T}.$$

For this class of weakly differentiable functions, we prove the following.

Lemma 2. Let $u \in H^s(\mathbb{R}^2)$, and let $f : (0, T = 2\pi R) \rightarrow \mathbb{R}$ be its trace on a circle of radius $R > 0$. Define the “ N -points approximation” of its average $\hat{f}(0)$ as

$$\hat{f}_N(0) := \frac{1}{N} \sum_{j=0}^{N-1} f(j \frac{T}{N}), \quad \hat{f}(0) = \frac{1}{T} \int_0^T f(x) \, dx = \frac{1}{2\pi} \int_0^{2\pi} f(\theta) \, d\theta, \quad \theta = \frac{x}{R};$$

then either

- the method is exact, $\hat{f}(0) = \hat{f}_N(0)$ if $\hat{f}(k) \equiv 0$ when $|k| \geq N$, or
- its “aliasing error” satisfies

$$(9) \quad \boxed{|\hat{f}(0) - \hat{f}_N(0)| \leq \sum_{k \in \mathbb{Z}_*} |\hat{f}(Nk)| \leq O(R^{\min(N, s - \frac{1}{2})}).}$$

Proof. See Appendix A. □

Proposition 1. Assume the “exact fluxes” $\mathcal{I}_{i,j}(\theta)$ are known for $\theta \in (0, 2\pi)$; then, for $u \in H^s(\mathbb{R}^2)$, the scheme (8) is endowed with a Local Truncation Error $O(R^{\min(3, s - \frac{5}{2})})$. If, instead, $u \in C^s(\mathbb{R}^2)$, it improves to $O(R^{\min(3, s - 2)})$.

Proposition 1 means that, as soon as the exact solution to (1) is smooth enough, approximating the circular integral in (6) with a “4-points trapezoidal rule” in (7) isn’t very costly in terms of numerical accuracy.

Proof. The first statement is a consequence of standard trace theory for 2D Sobolev functions on circles; see [29]. Since, for any circle C_R , the Besov space $B_{p,p}^{1-\frac{1}{p}}(C_R) = H^{\frac{1}{2}}(C_R)$ when $p = 2$, we immediately deduce from [29, Theorem 1] that the trace on $C_R(i, j)$ belongs to $H^{s-\frac{1}{2}}$, so the integrand in (6) is $H^{s-\frac{3}{2}}$. Lemma 2 yields that the “method of rectangles” is either exact, or of accuracy $O(R^{\min(4, s - \frac{3}{2})})$. There is still a factor $1/2R$ in (7) which yields accuracy $O(R^{\min(3, s - \frac{5}{2})})$. When $u \in C^s(\mathbb{R}^2)$, its restriction to C_R is endowed with the same regularity; the result easily follows. □

Remark 1. It isn’t easy to directly estimate the (curvilinear) derivatives of the trace f with respect to the abscissa x because of the circle’s curvature, which is $\frac{1}{R}$:

$$\begin{aligned} \frac{df}{dx} &= \frac{1}{R} \frac{df}{d\theta} = -\partial_x u(R \cos \theta, R \sin \theta) \sin \theta + \partial_y u(R \cos \theta, R \sin \theta) \cos \theta, \\ \frac{d^2 f}{dx^2} &= -\frac{1}{R} \left(\partial_x u(R \cos \theta, R \sin \theta) \cos \theta + \partial_y u(R \cos \theta, R \sin \theta) \sin \theta \right) + \dots, \end{aligned}$$

hence a “bad term” $\frac{1}{R} \frac{\partial u}{\partial r}$ already appears in the second derivative of f .

Elementary trace estimates on a circle are recalled in Appendix C.

2.3. Numerical fluxes for the “2D Bessel scheme”. Since even (7) is impractical, the scheme derived in [10] proceeds by locally inverting the stationary operator in (1) in order to produce “reliable values of both $\frac{\partial u}{\partial r}(R, \theta)$ and $u(R, \theta)$ at $\theta_k = k\pi/2$ ”. Indeed, consider the four disks $D_R(i \pm \frac{1}{2}, j \pm \frac{1}{2})$ (see Figure 2) the velocities $\mathbf{V}_{i \pm \frac{1}{2}, j \pm \frac{1}{2}}$ at their centers are exactly the ones required in (7):

- Given a T -periodic trace $g(x)$ with $T = 2\pi R$ on the boundary of each disk, it induces a solution $v(r, \theta)$, $r \leq R$, $\theta \in (0, 2\pi)$, of the stationary problem,

$$-\nabla \cdot (\varepsilon \nabla v - v \mathbf{V}) = 0, \quad v = g \text{ on } C_R := \partial D_R,$$

that is expressed thanks to its Dirichlet-Green function, G_ε^0 :

$$v(r, \theta) = -\varepsilon \int_0^{2\pi} h(\psi) \frac{\partial G_\varepsilon^0}{\partial \rho}(r, \theta; R, \psi) R \, d\psi,$$

where the function $h(\psi) = g(R\psi)$ is now 2π -periodic; see [10, eq. (2.8-9)].

- Accordingly, at their centers $(x_{i\pm\frac{1}{2},j\pm\frac{1}{2}})$, seemingly different numerical fluxes,

$$\mathcal{F}_{i,j}(\theta) := \left(\varepsilon \frac{\partial v}{\partial r} - v \|\mathbf{V}\| \cos(\mu - \theta) \right) (r = 0, \theta),$$

can be computed by the same procedure, because (see also [10, Prop. 2])

$$\frac{\partial v}{\partial r}(r, \theta) = -\varepsilon \int_0^{2\pi} h(\psi) \frac{\partial^2 G_\varepsilon^0}{\partial r \partial \rho}(r, \theta; R, \psi) R \, d\psi.$$

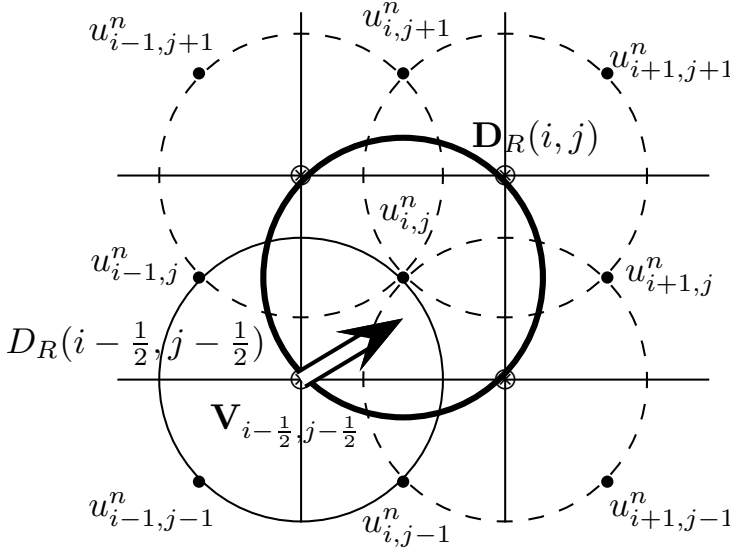


FIGURE 2. Schematic view of “Bessel scheme” on a uniform grid.

Lemma 3 (see [10]). *Let $\theta \in (-\pi, \pi)$ and $E(\psi) = \exp(-\omega_\varepsilon R \cos(\psi - \mu))$,*

$$(10) \quad \mathcal{F}_{i,j}(\theta) = \omega \int_0^{2\pi} h(\psi) E(\psi) \left(-\frac{\cos(\mu - \theta)}{I_0(\omega_\varepsilon R)} + \frac{\cos(\theta - \psi)}{I_1(\omega_\varepsilon R)} \right) \frac{d\psi}{2\pi},$$

where $\omega = \|\mathbf{V}\|/2$, $\omega_\varepsilon = \omega/\varepsilon$ and $I_n(\cdot)$ are modified Bessel functions of order n .

Proof. One gets from [10] that, for any of the four disks $D_R(i \pm \frac{1}{2}, j \pm \frac{1}{2})$,

$$\begin{aligned} v(0, \theta) &= \frac{1}{I_0(\omega_\varepsilon R)} \int_0^{2\pi} h(\psi) E(\psi) \frac{d\psi}{2\pi}, \\ \varepsilon \frac{\partial v}{\partial r}(0, \theta) &= \omega \cos(\theta - \mu) v(0, \theta) + \frac{\omega}{I_1(\omega_\varepsilon R)} \int_0^{2\pi} h(\psi) E(\psi) \cos(\theta - \psi) \frac{d\psi}{2\pi}. \end{aligned}$$

Since $\|\mathbf{V}\| \cos(\mu - \theta) = 2\omega \cos(\mu - \theta)$, diffusive and transport terms yield

$$\begin{aligned} D^\mu(\theta) - T^\mu(\theta) &= \varepsilon \frac{\partial v}{\partial r}(0, \theta) - \|\mathbf{V}\| \cos(\mu - \theta) v(0, \theta) \\ &= -\omega \cos(\theta - \mu) v(0, \theta) + \frac{\omega}{I_1(\omega_\varepsilon R)} \int_0^{2\pi} h(\psi) E(\psi) \cos(\theta - \psi) \frac{d\psi}{2\pi} \\ &= -\frac{\omega \cos(\mu - \theta)}{I_0(\omega_\varepsilon R)} \int_0^{2\pi} h(\psi) E(\psi) \frac{d\psi}{2\pi} + \frac{\omega}{I_1(\omega_\varepsilon R)} \int_0^{2\pi} h(\psi) E(\psi) \cos(\theta - \psi) \frac{d\psi}{2\pi}. \end{aligned}$$

□

2.4. Trigonometric interpolation and aliasing errors. According to [10, §3.1], the circular integrals used in Lemma 3 must be approximated by means of a trigonometric polynomial interpolating the four grid points lying on each circle,

$$(11) \quad h_4(\psi) = \mathbf{a}_0 + \mathbf{a}_1 \cos \psi + \mathbf{b}_1 \sin \psi + \mathbf{a}_2 \cos 2\psi, \quad \psi \in (0, 2\pi),$$

so that, for instance with $C_R(i - \frac{1}{2}, j - \frac{1}{2})$,

$$(12) \quad h_4(0) = u_{i,j}^n, \quad h_4\left(\frac{\pi}{2}\right) = u_{i-1,j}^n, \quad h_4(\pi) = u_{i-1,j-1}^n, \quad h_4\left(\frac{3\pi}{2}\right) = u_{i,j-1}^n.$$

It's not difficult to see that the values of the coefficients in (11) obtained by inverting the 4×4 linear system are identical to the ones one would obtain by applying the “4-points method of rectangles” on the exact function $h(\psi)$. Accordingly, define

$$(13) \quad \tilde{g}(x) = \sum_{-\frac{N}{2} < k \leq \frac{N}{2}} \tilde{g}_k \exp(ik \frac{2\pi x}{T}), \quad \tilde{g}_k = \frac{1}{N} \sum_{\ell=0}^{N-1} h\left(\frac{2\pi \ell}{N}\right) \exp(-ik \frac{2\pi \ell}{N});$$

then \tilde{g} is the N -points interpolation of h from the (unitary) matrix of the Discrete Fourier Transform, which inverse (transpose) matches the “method of rectangles”:

$$\forall \psi \in (0, 2\pi), \quad h_4(\psi) = \tilde{g}(\psi R) \text{ for } N = 4.$$

Indeed, as $\sin(k\pi) \equiv 0$ for $k \in \{0, 1, 2, 3\}$, the matrix of (11)–(12) satisfies

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix}, \quad M^{-1} = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{2} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & -\frac{1}{2} & -\frac{1}{4} \end{pmatrix},$$

so that $\mathbf{a}_0, \mathbf{a}_1, \mathbf{b}_1$ agree with \tilde{g}_k , $k \in \{0, \pm 1\}$ for $N = 4$. As $\exp(ik\pi) = \cos(k\pi)$ for $k \in \mathbb{N}$, the “Nyquist term” \tilde{g}_2 in (13) is equal to \mathbf{a}_2 , obtained from matrix inversion,

$$\tilde{g}(x) = \tilde{g}_{-1} \exp(-i \frac{x}{R}) + \tilde{g}_0 + \tilde{g}_1 \exp(i \frac{x}{R}) + \tilde{g}_2 \exp(2i \frac{x}{R}) = h_4\left(\frac{x}{R}\right).$$

Lemma 4. Let $u \in H^s(\Omega)$, $s \geq N + \frac{1}{2}$ with $T = 2\pi R$ and \tilde{g} be given by (13); then

$$\frac{1}{T} \int_0^T |g(x) - \tilde{g}(x)|^2 dx \leq O(R^N) \|u\|_{H^s}^2, \quad N \in 2\mathbb{N}, \quad x = R\psi.$$

Proof. Being T -periodic, the function g rewrites as a Fourier series,

$$g(x) = \sum_{k \in \mathbb{Z}} \hat{g}(k) \exp(ik \frac{x}{R}), \quad \hat{g}(k) = \int_0^T g(x) \exp(-ik \frac{x}{R}) \frac{dx}{T}.$$

The discrepancy between \tilde{g} and g is twofold: a lack of high-frequency coefficients and errors on low frequencies. Accordingly, we treat them in two distinct steps:

$$\|g - \tilde{g}\|_{L^2} \leq \underbrace{\|g - g_N\|_{L^2}}_{\text{high frequencies}} + \underbrace{\|g_N - \tilde{g}\|_{L^2}}_{\text{low frequencies}}.$$

- Let $N \in 2\mathbb{N}$ be an even integer, and define the truncated Fourier series,

$$(0, T) \ni x \mapsto g_N(x) = \sum_{-\frac{N}{2} < k \leq \frac{N}{2}} \hat{g}(k) \exp(ik \frac{x}{R}).$$

Taking advantage of (21) in Parseval's identity gives:

$$\begin{aligned} \frac{1}{T} \int_0^T |g(x) - g_N(x)|^2 dx &= \sum_{-\frac{N}{2}, |k| > \frac{N}{2}} |\hat{g}(k)|^2 \\ &\leq \sum_{-\frac{N}{2}, |k| > \frac{N}{2}} O(R^{2|k|}) \|u\|_{H^s}^2 \\ &\leq O(R^N) \quad \text{because } s \geq \frac{N+1}{2}. \end{aligned}$$

- Yet, errors on the low-frequency coefficients are given by Lemma 2:

$$(14) \quad \left| \hat{g}(k) - \frac{1}{N} \sum_{\ell=0}^{N-1} g\left(\frac{\ell T}{N}\right) \exp(-ik \frac{2\pi \ell}{N}) \right| \leq \sum_{\ell \in \mathbb{Z}^*} |\hat{g}(k + \ell N)|,$$

so that, again by Parseval's identity,

$$\begin{aligned} \frac{1}{T} \int_0^T |\tilde{g}(x) - g_N(x)|^2 dx &= \sum_{|k| < \frac{N}{2}} \left| \hat{g}(k) - \frac{1}{N} \sum_{\ell=0}^{N-1} g\left(\frac{\ell T}{N}\right) \exp(-ik \frac{2\pi \ell}{N}) \right|^2 \\ &\leq \sum_{|k| < \frac{N}{2}} \left(\sum_{\ell \in \mathbb{Z}^*} |\hat{g}(k + \ell N)|^2 \right). \end{aligned}$$

We again take advantage of (21) to bound each term for $k = 0, \pm 1, \dots, \frac{N}{2}$:

- for $k = 0$, the right-hand side of (14) is symmetric, so

$$\sum_{\ell \in \mathbb{Z}^*} |\hat{g}(k + \ell N)|^2 = 2 \sum_{\ell \in \mathbb{N}^*} |\hat{g}(\ell N)|^2 \leq O(R^{2N}), \quad \hat{g}(-k) = \hat{g}(k)^*,$$

because g is a real function and $s \geq N + \frac{1}{2}$;

- for $k = 1$, the right-hand side of (14) isn't any more symmetric, so

$$\sum_{\ell \in \mathbb{Z}^*} |\hat{g}(1 + \ell N)|^2 = |\hat{g}(1 - N)|^2 + |\hat{g}(N + 1)|^2 + \sum_{|\ell| > 1} |\hat{g}(1 + \ell N)|^2 \leq O(R^{2(N-1)});$$

- the “Nyquist term” (or “Nyquist frequency”) $k = \frac{N}{2}$ yields:

$$\sum_{\ell \in \mathbb{Z}^*} \left| \hat{g}\left(\left(\ell + \frac{1}{2}\right)N\right) \right|^2 = \left| \hat{g}\left(-\frac{N}{2}\right) \right|^2 + 2 \sum_{\ell \in \mathbb{N}^*} \left| \hat{g}\left(\left(\ell + \frac{1}{2}\right)N\right) \right|^2 \leq O(R^N).$$

Both high- and low-frequency components yield the announced estimate. \square

The smoothness of u in Lemma 4 could be relaxed up to $s \geq \frac{N+1}{2}$ because the error is dominated by the “Nyquist term”, which generally is of order $R^{\frac{N}{2}}$. Hence, it would suffice to have $|\hat{g}(\pm N)| = O(R^{\frac{N}{2}})$ instead of $O(R^N)$. Since the circular trace of u would belong to $H^{\frac{N}{2}}(0, T)$, necessary bounds still follow from (21).

Remark 2. Defining the “Nyquist term” g_N with $\hat{g}(\frac{N}{2})$, instead of $\hat{g}(-\frac{N}{2})$, has no incidence in keeping Fourier coefficients \hat{g}_k like the ones written in [10, eq. (3.4)]

with $N = 4$, hence overall consistency with (1). Moreover, when $K = \frac{N}{2}$, (13) gives

$$\begin{aligned}\hat{h}(K) &= \frac{1}{2\pi} \int_0^{2\pi} h(\psi) \exp(-iK\psi) \, d\psi, \quad \psi = \frac{x}{R} \\ &\simeq \frac{1}{2K} \sum_{\ell=0}^{2K-1} h\left(\frac{\pi\ell}{K}\right) \exp(-i\pi\ell) \\ &\simeq \frac{1}{2K} \sum_{\ell=0}^{2K-1} h\left(\frac{\pi\ell}{K}\right) (-1)^\ell \\ &\simeq \frac{1}{2K} \sum_{\ell=0}^{2K-1} h\left(\frac{\pi\ell}{K}\right) \cos(\pi\ell) \simeq \frac{1}{2\pi} \int_0^{2\pi} h(\psi) \cos(K\psi) \, d\psi.\end{aligned}$$

Yet, for this particular frequency, the aliasing formula (20) gives

$$\tilde{h}_K - \hat{h}(K) = \hat{h}(-K) + \sum_{\ell \in 2\mathbb{N}^*+1} \hat{h}(K\ell) + \hat{h}(-K\ell),$$

so that, for the “real coefficient” $\hat{a}(K) = \hat{h}(K) + \hat{h}(-K)$, it comes

$$\tilde{h}_K - \hat{a}(K) = \sum_{\ell \in 2\mathbb{N}^*+1} \hat{h}(K\ell) + \hat{h}(-K\ell) \leq O(R^{3K}).$$

Lemma 4 furnishes essentially a bound on “aliasing errors”, a common use in Sampling Theory; see, e.g., [27, page 106]. This shouldn’t be a surprise because the approximation (11) of boundary data $h(\psi)$ is a “band-limited reconstruction” from four grid samples. This is a truly multi-dimensional feature, because it is not needed when deriving one-dimensional well-balanced schemes; see, e.g., [2, 17, 20].

2.5. What happens for steady-states? Given a (continuous) stationary regime $u \in H^s(\Omega)$, ideally, a (semi-discrete) scheme for (1) would be such that

$$\forall(i, j) \in \mathbb{Z}^2, \quad \frac{d \bar{u}_{i,j}(t)}{dt} \equiv 0.$$

However, the only thing we know for sure is that, for such initial data, the exact radial derivative written in (6) would be strictly identical to $\mathcal{F}_{i,j}(\theta)$ written in Lemma 3. Since the boundary data h must be truncated to get a computable expression, some (aliasing) error, bounded by Lemma 4, enters into the integral.

Lemma 5. *Let $u(x, y) \in H^{\frac{7}{2}}(\Omega)$ be a steady-state for (1), and let h_4 defined as in (11) be a “4-points trigonometric interpolation” of its circular trace, h . Then*

$$\left| \int_0^{2\pi} (h(\psi) - h_4(\psi)) E(\psi) \left[\frac{\omega \cos(\mu - \theta)}{I_0(\omega_\varepsilon R)} - \frac{\omega \cos(\theta - \psi)}{I_1(\omega_\varepsilon R)} \right] \frac{d\psi}{2\pi} \right| \leq O(R^2).$$

Proof. We again split the discrepancy into low and high frequencies by writing,

$$\begin{aligned}h(\psi) - h_4(\psi) &= (\hat{h}(0) - \mathbf{a}_0) + (\hat{a}(1) - \mathbf{a}_1) \cos \psi + (\hat{b}(1) - \mathbf{b}_1) \sin \psi \\ &\quad + (\hat{a}(2) - \mathbf{a}_2) \cos 2\psi + \hat{b}(2) \sin 2\psi + \underbrace{\sum_{k \geq 2} \hat{a}(k) \cos k\psi + \hat{b}(k) \sin k\psi}_{\text{high frequencies}}.\end{aligned}$$

- By Cauchy-Schwarz inequality, high frequencies are bounded by,

$$\left[\sum_{|k| \geq 3} |\hat{h}(k)|^2 \right]^{\frac{1}{2}} \left[\int_0^{2\pi} E(\psi)^2 \left[\frac{\omega \cos(\mu - \theta)}{I_0(\omega_\varepsilon R)} - \frac{\omega \cos(\theta - \psi)}{I_1(\omega_\varepsilon R)} \right]^2 \frac{d\psi}{2\pi} \right]^{\frac{1}{2}},$$

where the first term is $O(R^{\min(3, s - \frac{1}{2})})$ by (21). If $u \in H^{\frac{7}{2}}(\Omega)$, it has a trace in $H^3(0, T)$. To estimate the second term involving an exponential,

$$E(\psi)^2 = \exp(-2\omega_\varepsilon R \cos(\psi - \mu)),$$

thus, as $2ab \leq a^2 + b^2$, $(a - b)^2 \leq 2(a^2 + b^2)$ and

$$\left(\frac{\omega \cos(\mu - \theta)}{I_0(\omega_\varepsilon R)} - \frac{\omega \cos(\theta - \psi)}{I_1(\omega_\varepsilon R)} \right)^2 \leq 2\omega^2 \left(\frac{\cos^2(\mu - \theta)}{I_0(\omega_\varepsilon R)^2} + \frac{\cos^2(\theta - \psi)}{I_1(\omega_\varepsilon R)^2} \right),$$

we use the “integral representations of Bessel functions”; see [10, §3]:

$$\int_0^{2\pi} E(\psi)^2 \cos^2(\mu - \theta) \frac{d\psi}{2\pi} \leq \frac{1}{2\pi} \int_0^{2\pi} \exp(-2\omega_\varepsilon R \cos \psi) \frac{d\psi}{2\pi} = I_0(2\omega_\varepsilon R).$$

Accordingly, when $\omega_\varepsilon R \ll 1$, the second term is bounded by:

$$(15) \quad 2\omega \sqrt{2I_0(2\omega_\varepsilon R)} \left(\frac{1}{I_0(\omega_\varepsilon R)^2} + \frac{1}{I_1(\omega_\varepsilon R)^2} \right)^{\frac{1}{2}} \leq O\left(\frac{1}{R}\right).$$

Thus, errors resulting from the lack of high frequencies are $O(R^{\min(2, s - \frac{3}{2})})$.

- By linearity, low-frequency errors are retrieved by exactly computing the integral involving the 5 first Fourier terms: This can be done by following the calculations made in [10, §3.2–3]. Denote $\mathbf{e}_0 = \hat{h}(0) - \mathbf{a}_0$, $\mathbf{e}_1 = \hat{a}(1) - \mathbf{a}_1$, $\tilde{\mathbf{e}}_1 = \hat{b}(1) - \mathbf{b}_1$, $\mathbf{e}_2 = \hat{a}(2) - \mathbf{a}_2$; from (10) and [10, §4.1], it comes

$$\begin{aligned} E^\mu(\theta) = & \left\{ 2\mathbf{e}_0 - (\mathbf{e}_1 \cos \mu + \tilde{\mathbf{e}}_1 \sin \mu) \left(\frac{I_1(\omega_\varepsilon R)}{I_0(\omega_\varepsilon R)} + \frac{I_0(\omega_\varepsilon R)}{I_1(\omega_\varepsilon R)} - \frac{1}{\omega_\varepsilon R} \right) \right. \\ & \left. + \mathbf{e}_2 \cos(2\mu) \left(1 + \frac{I_2(\omega_\varepsilon R)}{I_0(\omega_\varepsilon R)} \right) \right\} \omega_\varepsilon \cos(\theta - \mu) + \frac{(\mathbf{e}_1 \sin \mu - \tilde{\mathbf{e}}_1 \cos \mu) \sin(\theta - \mu)}{R}. \end{aligned}$$

From this expression, it is clear that, when $\omega_\varepsilon R \ll 1$, both \mathbf{e}_0 and \mathbf{e}_2 are multiplied by $O(1)$ quantities, but \mathbf{e}_1 and $\tilde{\mathbf{e}}_1$ are multiplied by $O(\frac{1}{R})$. These terms are bounded by applying (20) and (21) with $s > \frac{1}{2}$, namely

$$|\mathbf{e}_0| = O(R^{\min(4, s - \frac{1}{2})}), \quad |\mathbf{e}_1|, |\tilde{\mathbf{e}}_1| = O(R^{\min(3, s - \frac{1}{2})}), \quad |\mathbf{e}_2| = O(R^{\min(6, s - \frac{1}{2})}).$$

Concerning the last $\hat{b}(2)$ term, by explicitly computing the integral,

$$\int_0^{2\pi} \sin 2\psi E(\psi) \left[\frac{\omega \cos(\mu - \theta)}{I_0(\omega_\varepsilon R)} - \frac{\omega \cos(\theta - \psi)}{I_1(\omega_\varepsilon R)} \right] \frac{d\psi}{2\pi},$$

following [10, page 2859], one finds that its contribution is again in $O(1)\hat{b}(2) = O(R^{\min(2, s - \frac{1}{2})})$ thanks to both (21) and the smoothness of u . All in all, errors resulting from low frequencies are $O(R^{\min(2, s - \frac{3}{2})})$.

□

Remark 3. Lemma 5 quantifies the error made when replacing exact fluxes $\mathcal{J}_{i,j}(\theta)$ by approximate ones $\mathcal{F}_{i,j}(\theta)$ resulting from a “4-points trapezoidal rule” on the circle of radius R . A similar statement holds when $u(x, y) \in H^s(\Omega)$, $s > \frac{3}{2}$:

$$(16) \quad |\mathcal{J}_{i,j}(\theta) - \mathcal{F}_{i,j}(\theta)| \leq O(R^{\min(2, s - \frac{3}{2})}).$$

This bound is compatible with the regularity theory of stationary solutions occurring in a Lipschitz convex domain like the unit square; see Lemma 1 and [8, 32].

It remains to prove Assertion (1) of the Main Theorem 1.

Proposition 2. *Pick, as initial data, a steady-state $u^0 \in H^s(\Omega)$, $s > \frac{3}{2}$ for (1); then the (initial) Local Truncation Error of the “2D Bessel scheme” of [10] satisfies,*

$$\boxed{\forall (i, j) \in \mathbb{Z}^2, \quad \begin{cases} \sum_{k=0}^3 \mathcal{F}_{i,j}(k\pi/2) = O(R^{\min(2, s - \frac{3}{2})}), & (\text{“2D WB fluxes”}) \\ |u_{i,j}(\Delta t) - u^0(x_i, y_j)| \leq O\left(R^{\min(3, s - \frac{1}{2}) - \chi(\Delta x \geq \varepsilon)}\right), \end{cases}}$$

Proof. It remains to gather all the intermediate estimates.

- Proposition 1 yields the “averaging error” for the circular integral in (6),

$$\begin{aligned} \frac{d u_{i,j}}{d t} &= \frac{1}{\pi R} \int_0^{2\pi} \left(\varepsilon \frac{\partial u}{\partial r}(R, \theta) - \|\mathbf{V}\| \cos(\mu - \theta) u(R, \theta) \right) d\theta \\ &= \frac{1}{2R} \sum_{k=0}^3 \mathcal{J}_{i,j}(k\pi/2) + O(R^{\min(3, s - \frac{5}{2})}); \end{aligned}$$

- Yet, there is also an error coming from using approximate fluxes $\mathcal{F}_{i,j}$ instead of exact ones $\mathcal{J}_{i,j}$, which is bounded by Lemma 5 and (16). Finally,

$$\frac{d u_{i,j}}{d t} = \frac{1}{2R} \sum_{k=0}^3 \mathcal{F}_{i,j}(k\pi/2) + O(R^{\min(1, s - \frac{5}{2})}) + O(R^{\min(3, s - \frac{5}{2})}),$$

and since the initial data $u^0 \in H^s(\Omega)$ is at steady-state, it comes

$$\frac{1}{2R} \sum_{k=0}^3 \mathcal{F}_{i,j}(k\pi/2) = O(R^{\min(1, s - \frac{5}{2})}) + O(R^{\min(3, s - \frac{5}{2})}).$$

Integrating this expression with respect to time, a factor $\Delta t/2R$ multiplies the numerical fluxes. In diffusive regime, when $\Delta x < \varepsilon$, the CFL condition $\Delta t/2R = O(R)$ gives the announced bound in $O(R^{\min(3, s - \frac{1}{2})})$. Oppositely, if $\Delta x \geq \varepsilon$ and transport dominates, $\Delta t/2R = O(1)$ so the accuracy decreases to $O(R^{\min(2, s - \frac{3}{2})})$. \square

2.6. Numerical tests. To assess our estimates, we set up (1) in the unit square,

$$\varepsilon = 1/100, \quad \mathbf{V} = \nabla^\perp \psi \text{ with } \psi(x, y) = -\sin(\pi x) \sin(\pi y)/\pi.$$

Boundary data was chosen as $\sin(\pi x)$, being a function in $C^\infty(0, 4)$, the perimeter of the computational domain. For $t \simeq 10$, the scheme in [10] reaches a stationary regime, and for the 96×96 grid, it was considered a “reference solution”, \bar{u} . Errors and convergence rates with respect to Δx were studied with the usual formula,

$$E_{p,N} = \|\bar{u} - u_N\|_{L^p} = \left(\Delta x^2 \|\bar{u} - u_N\|_{\ell^p}^p \right)^{\frac{1}{p}}, \quad r_{p,N} = \frac{\log(E_{p,N}/E_{p,M})}{\log(M/N)},$$

TABLE 1. Measured and “interpolated” convergence rates ($\varepsilon = \frac{1}{100}$).

Grids	L^1 rate	L^2 rate	“interpolated” L^2	L^4 rate
$N \in \{12, 16\}$	1.2612762	1.2442015	1.1661154	1.1212623
$N \in \{16, 24\}$	1.3234812	1.3142708	1.2901447	1.2737926
$N \in \{24, 32\}$	1.4283889	1.4499337	1.435143	1.438532
$N \in \{32, 48\}$	1.7231812	1.7446693	1.7354043	1.7415483

where N, M stand for different numbers of grid-points. Figure 3 indicates that:

- Convergence rates $r_{p,N}$ increase with N (i.e., when Δx decreases); five grids were compared to the reference solution, $N \in \{12, 16, 24, 32, 48\}$;
- Errors grow with p (as must be in a bounded domain), and rates tend to (slightly) decrease with p , especially when $\Delta x/\varepsilon$ becomes big.

This agrees with Proposition 2 which states that rates lower when transport dominates.

Remark 4 (Compatibility with interpolation inequalities). A classical inequality states that, for any f belonging to the convenient spaces,

$$(17) \quad \|f\|_{L^2} \leq \left(\|f\|_{L^p}\right)^{1-\theta} \left(\|f\|_{L^q}\right)^{\theta}, \quad \frac{1}{2} = \frac{1-\theta}{p} + \frac{\theta}{q}.$$

By applying it to $p = 1$, $q = 4$ with $\theta = 2/3$, we can check whether the convergence rates that were observed numerically turn out to be compatible with (17). Indeed, for a given $n \times n$ grid with $M \leq n \leq N$, based on (17), we can expect that

$$E_{2,n} \leq (E_{1,n})^{\frac{1}{3}} (E_{4,n})^{\frac{2}{3}} \quad \text{with} \quad E_{p,n} \simeq O(\Delta x^{r_{p,n}}).$$

Yet, for the grids which were used in the test, $n = N$, so that (17) yields

$$\boxed{\forall N \in \{12, 16, 24, 32, 48\}, \quad r_{2,N} \geq r_{1,N}^{\frac{1}{3}} \cdot r_{4,N}^{\frac{2}{3}}.}$$

Accordingly, Table 1 reveals a general agreement between the “interpolated L^2 rates” retrieved by (17) and the ones measured for $\varepsilon = \frac{1}{100}$, displayed in Figure 3.

For completeness, the numerical test was carried out with both $\varepsilon = \frac{1}{400}$ and $\varepsilon = \frac{1}{1000}$ on identical computational grids (see Figure 3). Sharper boundary layers yield bigger errors, the map $p \mapsto E_{p,N}$ increases, as expected in a bounded domain.

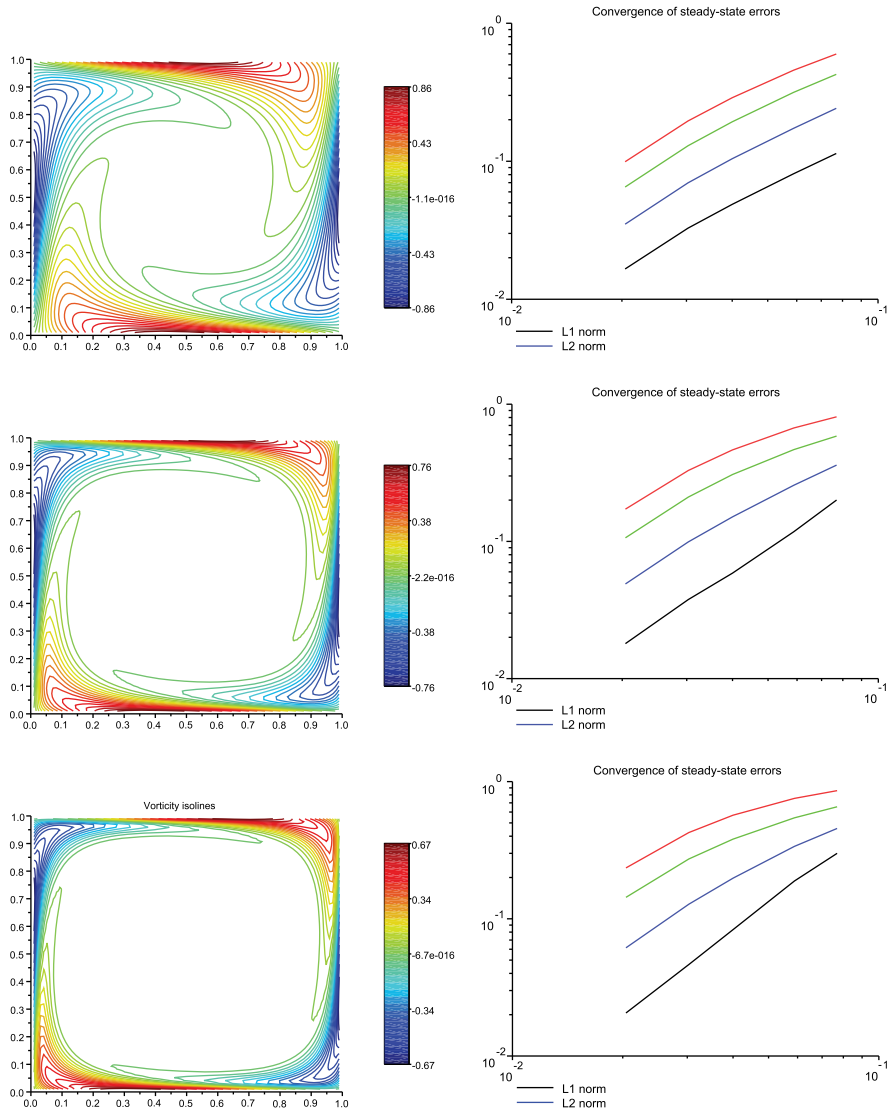


FIGURE 3. Steady-states (left), convergence in L^p , $p \in \{1, 2, 4, 8\}$ (right) for viscosity coefficients $\varepsilon = \frac{1}{100}, \frac{1}{400}, \frac{1}{1000}$ (top to bottom).

3. AN L^2 STABILITY FOR THE SEMI-DISCRETE “BESSEL SCHEME”

Hereafter, we prove Assertions (2), (3), and (4) in the Main Theorem 1.

3.1. Elementary 1D Scharfetter-Gummel scheme. As a preliminary step, a divergence-free vector field in 1D is constant, so (1) reduces to the simple equation,

$$\partial_t u + a \partial_x u = \varepsilon \partial_{xx} u, \quad a \in \mathbb{R}.$$

The (uniformly accurate) Scharfetter-Gummel scheme (see, e.g., [15, 32, 34] and [10, Appendix A]) is stable at all Peclet numbers $|a|\Delta x/2\varepsilon$. Let $k = |a|/2\varepsilon$; it reads:

$$\frac{d u_j(t^n)}{dt} \simeq \frac{u_j^{n+1} - u_j^n}{\Delta t} = \frac{\varepsilon k}{\Delta x} \left\{ \frac{e^{-k\Delta x} u_{j+1} - e^{k\Delta x} u_j}{\sinh k\Delta x} - \frac{e^{-k\Delta x} u_j - e^{k\Delta x} u_{j-1}}{\sinh k\Delta x} \right\}.$$

We multiply and then sum on $j \in \mathbb{Z}$ in order to get a negative sign:

$$\begin{aligned} \frac{d}{dt} \left(\sum_j \frac{|u_j|^2}{2} \right) &= -\frac{\varepsilon k}{\Delta x} \sum_j (u_{j-1} - u_j) \quad (\text{“summation by parts”}) \\ &\quad \times \frac{u_j(\sinh - \cosh)(k\Delta x) + u_{j-1}(\sinh + \cosh)(k\Delta x)}{\sinh(k\Delta x)} \\ &= -\frac{\varepsilon k}{\Delta x} \sum_j (u_j^2 - u_{j-1}^2) + \frac{1}{\tanh(k\Delta x)} |u_j - u_{j-1}|^2 \\ &= -\varepsilon \sum_j \frac{k |u_j - u_{j-1}|^2}{\Delta x \tanh(k\Delta x)} \leq 0. \end{aligned}$$

This easy computation reveals two important things:

- Centered differences are well-suited for deriving “telescopic sums” which, in turn, allows us to cancel contributions coming from transport terms;
- Hence one has to split the scheme into a “centered part” (meant to be telescopic) and an “artificial viscosity” part (which brings a negative sign).

3.2. The 2D “Bessel scheme” and its quadratic form. On any uniform Cartesian grid with parameter $\Delta x = \Delta y > 0$, along with $R = \Delta x/\sqrt{2}$, the (semi-discrete) “Bessel scheme” introduced in [10, §3–4] writes,

$$\begin{aligned} \forall (i, j) \in \mathbb{Z}^2, \quad \frac{d u_{i,j}(t^n)}{dt} &\simeq \frac{u_{i,j}^{n+1} - u_{i,j}^n}{\Delta t} \\ &= -\frac{\varepsilon}{2R} \left(F_{i-\frac{1}{2}, j-\frac{1}{2}}(0) + F_{i+\frac{1}{2}, j+\frac{1}{2}}(\pi) + F_{i+\frac{1}{2}, j-\frac{1}{2}}\left(\frac{\pi}{2}\right) + F_{i-\frac{1}{2}, j+\frac{1}{2}}\left(-\frac{\pi}{2}\right) \right), \end{aligned}$$

where “4-points numerical fluxes” read, for generic values $a, b, c, d \in \mathbb{R}^4$,

$$\begin{aligned} F^\mu(\theta) &:= a \mathcal{C}_a^\mu(\theta) + b \mathcal{C}_b^\mu(\theta) + c \mathcal{C}_c^\mu(\theta) + d \mathcal{C}_d^\mu(\theta) := \frac{\omega_\varepsilon \cos(\theta - \mu)}{2} \\ &\quad \times \left[a \left\{ 1 - \sin \mu \mathcal{B}_1(\omega_\varepsilon R) - \frac{\cos(2\mu)}{2} \mathcal{B}_2(\omega_\varepsilon R) - \frac{\cos \mu \tan(\theta - \mu)}{\omega_\varepsilon R} \right\} \right. \\ &\quad + b \left\{ 1 + \sin \mu \mathcal{B}_1(\omega_\varepsilon R) - \frac{\cos(2\mu)}{2} \mathcal{B}_2(\omega_\varepsilon R) + \frac{\cos \mu \tan(\theta - \mu)}{\omega_\varepsilon R} \right\} \\ &\quad + c \left\{ 1 - \cos \mu \mathcal{B}_1(\omega_\varepsilon R) + \frac{\cos(2\mu)}{2} \mathcal{B}_2(\omega_\varepsilon R) + \frac{\sin \mu \tan(\theta - \mu)}{\omega_\varepsilon R} \right\} \\ &\quad \left. + d \left\{ 1 + \cos \mu \mathcal{B}_1(\omega_\varepsilon R) + \frac{\cos(2\mu)}{2} \mathcal{B}_2(\omega_\varepsilon R) - \frac{\sin \mu \tan(\theta - \mu)}{\omega_\varepsilon R} \right\} \right], \end{aligned}$$

and $0 \leq \omega_\varepsilon = \frac{\|\mathbf{V}\|}{2\varepsilon}$ (the Euclidean \mathbb{R}^2 norm) along with

$$\mathcal{B}_1(\omega_\varepsilon R) = \left(\frac{I_1}{I_0} + \frac{I_0}{I_1} \right) (\omega_\varepsilon R) - \frac{1}{\omega_\varepsilon R}, \quad \mathcal{B}_2(\omega_\varepsilon R) = 1 + \frac{I_2}{I_0} (\omega_\varepsilon R).$$

For instance, in the (south-west) disk $D_R(i - \frac{1}{2}, j - \frac{1}{2})$, (see Figure 2)

$$(18) \quad a = u_{i-1,j}, \quad b = u_{i,j-1}, \quad c = u_{i,j}, \quad d = u_{i-1,j-1}.$$

Lemma 6. *Assume that the flux in (1) vanishes at infinity; then*

$$\forall \varepsilon, t > 0, \quad \frac{d}{dt} \left(\sum_{(i,j) \in \mathbb{Z}^2} u_{i,j}(t) \right) \equiv 0.$$

Proof. By shifting indexes in the summation, one derives,

$$\begin{aligned} \frac{d}{dt} \left(\sum_{i,j} u_{i,j} \right) &= -\frac{\varepsilon}{2R} \sum_{i,j} F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu(0) + F_{i+\frac{1}{2},j+\frac{1}{2}}^\mu(\pi) + F_{i+\frac{1}{2},j-\frac{1}{2}}^\mu\left(\frac{\pi}{2}\right) + F_{i-\frac{1}{2},j+\frac{1}{2}}^\mu\left(-\frac{\pi}{2}\right) \\ &= -\frac{\varepsilon}{2R} \sum_{i,j} F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu(0) + F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu(\pi) + F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu\left(\frac{\pi}{2}\right) + F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu\left(-\frac{\pi}{2}\right). \end{aligned}$$

Accordingly, in generic variables (18), the claim reduces to

$$F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu(0) + F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu(\pi) + F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu\left(\frac{\pi}{2}\right) + F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu\left(-\frac{\pi}{2}\right) = 0,$$

which is easy to verify, based on the former expression of $\theta \mapsto F^\mu(\theta)$. \square

Remark 5. Hereafter, it's convenient to work with slightly different notation compared to what was used in the former section. The reason is simple: to establish “2D well-balanced” properties, one must scrutinize $\frac{d u_{i,j}}{dt}$, which is a quantity lying at the “center” (i, j) . However, the L^2 estimate is retrieved by studying dissipation and cancellations inside a disk centered at a “node” $(i - \frac{1}{2}, j - \frac{1}{2})$. This is why we work now with $F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu(\theta)$ instead of $\mathcal{F}_{i,j}(\theta)$; indeed, $\mathcal{F}_{i,j}(-\pi) = F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu(0)$, $\mathcal{F}_{i,j}(\pi/2) = F_{i-\frac{1}{2},j+\frac{1}{2}}^\mu(-\pi/2)$, and so on.

As before, we multiply by $u_{i,j}(t)$ and sum on $i, j \in \mathbb{Z}^2$, so that

$$\begin{aligned} \frac{d}{dt} \left(\sum_{i,j} \frac{|u_{i,j}|^2}{2} \right) &= -\frac{\varepsilon}{2R} \sum_{i,j} u_{i,j} \left(F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu(0) + F_{i+\frac{1}{2},j+\frac{1}{2}}^\mu(\pi) + F_{i+\frac{1}{2},j-\frac{1}{2}}^\mu\left(\frac{\pi}{2}\right) + F_{i-\frac{1}{2},j+\frac{1}{2}}^\mu\left(-\frac{\pi}{2}\right) \right) \\ &= -\frac{\varepsilon}{2R} \sum_{i,j} F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu(0) u_{i,j} + F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu(\pi) u_{i-1,j-1} + F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu\left(\frac{\pi}{2}\right) u_{i-1,j} \\ &\quad + F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu\left(-\frac{\pi}{2}\right) u_{i,j-1}. \end{aligned}$$

The last line follows by “shifting indexes” in the summation, and highlights that an important quantity, when expressed in “generic variables” reads:

$$(19) \quad \boxed{c F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu(0) + d F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu(\pi) + a F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu\left(\frac{\pi}{2}\right) + b F_{i-\frac{1}{2},j-\frac{1}{2}}^\mu\left(-\frac{\pi}{2}\right)},$$

which is actually a quadratic form which spectral decomposition reveals is a positive eigenvalue; see Figure 4. Using shorthand notation $\mathcal{B}_1, \mathcal{B}_2$ for $\mathcal{B}_1(\omega_\varepsilon R), \mathcal{B}_2(\omega_\varepsilon R)$,

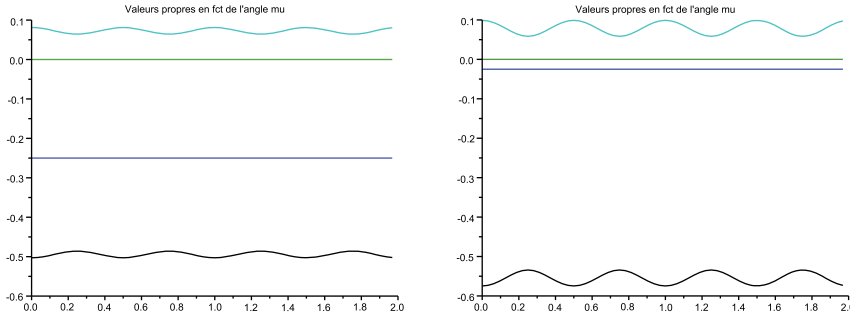


FIGURE 4. Numerical eigenvalues of the quadratic form (19) as functions of the angle μ for $\varepsilon = 1/100$ (left) and $\varepsilon = 1/1000$ (right).

for any $\omega_\varepsilon \in \mathbb{R}_*^+$ and $\mu \in (0, 2\pi)$, we rewrite two-dimensional “Bessel fluxes” like:

- For $\theta = 0$ or π :

$$\begin{aligned} F^\mu(0/\pi) = & \pm \frac{\omega_\varepsilon \cos \mu}{2} \left[(a+b) \left(1 - \frac{\mathcal{B}_2}{2} \cos 2\mu \right) + (c+d) \left(1 + \frac{\mathcal{B}_2}{2} \cos 2\mu \right) \right] \\ & \pm \frac{\sin \mu}{2R} \left[(a-b) \cos \mu - (c-d) \sin \mu \right] \quad (\text{“diffusive terms”}) \\ & \mp \frac{\omega_\varepsilon \cos \mu}{2} \mathcal{B}_1 \left[(a-b) \sin \mu + (c-d) \cos \mu \right] \quad (\text{“mixed terms”}), \end{aligned}$$

- For $\theta = \pi/2$ or $-\pi/2$:

$$\begin{aligned} F^\mu(\pm \frac{\pi}{2}) = & \pm \frac{\omega_\varepsilon \sin \mu}{2} \left[(a+b) \left(1 - \frac{\mathcal{B}_2}{2} \cos 2\mu \right) + (c+d) \left(1 + \frac{\mathcal{B}_2}{2} \cos 2\mu \right) \right] \\ & \mp \frac{\cos \mu}{2R} \left[(a-b) \cos \mu - (c-d) \sin \mu \right] \quad (\text{“diffusive terms”}) \\ & \mp \frac{\omega_\varepsilon \sin \mu}{2} \mathcal{B}_1 \left[(a-b) \sin \mu + (c-d) \cos \mu \right] \quad (\text{“mixed terms”}). \end{aligned}$$

Clearly, we see “centered differences” on the first line, and “diffusion terms” on the two next ones. However, the terms multiplied by \mathcal{B}_1 aren’t always of diffusive type because of the function \mathcal{B}_1 , which behaves like $1/(\omega_\varepsilon R)$ for small Peclet, but tends to 2 when the Peclet diverges to $+\infty$. Both \mathcal{B}_1 and \mathcal{B}_2 are positive, and

$$0 \leq \mathcal{B}_1, \mathcal{B}_2 \rightarrow 2, \quad \text{when } \omega_\varepsilon R \rightarrow +\infty.$$

3.3. The case of constant velocities (advection-diffusion). Yet, we need to treat each type of term differently: “transport terms” are meant to produce a telescopic sum (because $a+b$ and $c+d$ hint of the presence of centered differences), whereas “diffusion and mixed terms” should have a definite (negative) sign.

- Transport terms:

$$\begin{aligned} cF(0) + dF(\pi) \rightsquigarrow & \frac{\omega_\varepsilon \cos \mu}{2} \left[(a+b)(c-d) \left(1 - \frac{\mathcal{B}_2}{2} \cos 2\mu \right) \right. \\ & \left. + (c^2 - d^2) \left(1 + \frac{\mathcal{B}_2}{2} \cos 2\mu \right) \right], \end{aligned}$$

$$aF\left(\frac{\pi}{2}\right) + bF\left(-\frac{\pi}{2}\right) \rightsquigarrow \frac{\omega_\varepsilon \sin \mu}{2} \left[(a^2 - b^2) \left(1 - \frac{\mathcal{B}_2}{2} \cos 2\mu \right) + (c + d)(a - b) \left(1 + \frac{\mathcal{B}_2}{2} \cos 2\mu \right) \right],$$

and we claim that all the terms involved are telescopic (in particular, $2 \pm \mathcal{B}_2 \cos 2\mu \geq 0$ uniformly). It is obvious for both $c^2 - d^2$ and $a^2 - b^2$, but also for the two other ones which read:

$$(a + b)(c - d) = c(a + b) - d(a + b), \quad (c + d)(a - b) = a(c + d) - b(c + d).$$

Let's pick (for instance) the south-west square $i - \frac{1}{2}, j - \frac{1}{2}$,

$$\begin{aligned} \sum_{i,j} c(a + b) - d(a + b) &= \sum_{i,j} u_{i,j}(u_{i-1,j} + u_{i,j-1}) - \sum_{i,j} u_{i-1,j-1}(u_{i-1,j} + u_{i,j-1}) \\ &= \sum_{i,j} u_{i,j}(u_{i-1,j} + u_{i,j-1}) - \sum_{i,j} u_{i,j-1}u_{i,j} - \sum_{i,j} u_{i-1,j}u_{i,j} \\ &= 0. \end{aligned}$$

Hence, assuming that both ω_ε and μ are constants in \mathbb{R}^2 , the overall (telescopic) sum of “transport terms” vanishes.

- Diffusive terms:

$$\begin{aligned} cF(0) + dF(\pi) &\rightsquigarrow \frac{\sin \mu}{2R} \left[(a - b)(c - d) \cos \mu - (c - d)^2 \sin \mu \right], \\ aF\left(\frac{\pi}{2}\right) + bF\left(-\frac{\pi}{2}\right) &\rightsquigarrow \frac{\cos \mu}{2R} \left[-(a - b)^2 \cos \mu + (a - b)(c - d) \sin \mu \right]. \end{aligned}$$

The elementary inequality $2ab \leq a^2 + b^2$ yields

$$\boxed{2[(a - b) \cos \mu] \times [(c - d) \sin \mu] \leq |a - b|^2 \cos^2 \mu + |c - d|^2 \sin^2 \mu,}$$

so that the sum of these “diffusive terms” is nonpositive (and not better: an exponential decay from a Poincaré inequality isn't clear).

- Mixed terms:

$$\begin{aligned} cF(0) + dF(\pi) &\rightsquigarrow -\frac{\omega_\varepsilon \mathcal{B}_1 \cos \mu}{2} \left[(a - b)(c - d) \sin \mu + (c - d)^2 \cos \mu \right], \\ aF\left(\frac{\pi}{2}\right) + bF\left(-\frac{\pi}{2}\right) &\rightsquigarrow -\frac{\omega_\varepsilon \mathcal{B}_1 \sin \mu}{2} \left[(a - b)^2 \sin \mu + (a - b)(c - d) \cos \mu \right], \end{aligned}$$

and they are exactly like the “diffusive ones” before because $\mathcal{B}_1 \geq 0$.

All in all, we get the following (negative) quantity:

$$\begin{aligned} &\left(\frac{1}{R} - \omega_\varepsilon \mathcal{B}_1 \right) (a - b)(c - d) \cos \mu \sin \mu \\ &\quad - |a - b|^2 \left(\frac{\cos^2 \mu}{2R} + \frac{\omega_\varepsilon \mathcal{B}_1 \sin^2 \mu}{2} \right) - |c - d|^2 \left(\frac{\sin^2 \mu}{2R} + \frac{\omega_\varepsilon \mathcal{B}_1 \cos^2 \mu}{2} \right), \end{aligned}$$

where

$$\omega_\varepsilon \mathcal{B}_1 \geq 0, \quad \text{but} \quad \frac{1}{R} - \omega_\varepsilon \mathcal{B}_1 = -\omega_\varepsilon \left(\frac{I_1}{I_0} + \frac{I_0}{I_1} \right) \leq 0.$$

Proposition 3. *If $\mathbf{V} \in \mathbb{R}^2$ is a constant vector, then the “semi-discrete Bessel approximation” $t \mapsto u_{i,j}(t)$, dissipates the L^2 norm (or the “enstrophy”):*

$$\forall \varepsilon > 0, (i, j) \in \mathbb{Z}^2, \quad \frac{d}{dt} \sum_{i,j} |u_{i,j}(t)|^2 \leq 0.$$

Remark 6. Since the convection-diffusion equation with a constant velocity $\mathbf{V} \in \mathbb{R}^2$ is translation-invariant, and the “Bessel scheme” is linear in all its entries, this L^2 bound can easily be extended into an H^1 bound as soon as the initial data $u_0(x, y) \in H^1(\mathbb{R}^2)$. This isn’t true anymore if $\mathbf{V}(x, y)$ varies in (x, y) .

3.4. General case of divergence-free velocity fields. When $\mathbf{V}(x, y)$ varies in space, but is endowed with a null divergence, it is expected that only a part of the transport terms still can generate “telescopic sums” because of upwinding. In order to ease notation, let’s define:

$$\forall i, j \in \mathbb{Z}^2, \quad (\tilde{\mathcal{B}}_2)_{i-\frac{1}{2}, j-\frac{1}{2}} := \left(\mathcal{B}_2(\omega_\varepsilon R) \frac{\cos 2\mu}{2} \right)_{i-\frac{1}{2}, j-\frac{1}{2}} \in [-1, 1],$$

where now, $\mu = \mu_{i-\frac{1}{2}, j-\frac{1}{2}}$ and $\omega = \omega_{i-\frac{1}{2}, j-\frac{1}{2}}$ depend on i, j , so that,

$$\begin{aligned} cF(0) + dF(\pi) &\rightsquigarrow \frac{\omega_\varepsilon \cos \mu}{2} \left[(a+b)(c-d) \left(1 - \tilde{\mathcal{B}}_2 \right) + (c^2 - d^2) \left(1 + \tilde{\mathcal{B}}_2 \right) \right], \\ aF\left(\frac{\pi}{2}\right) + bF\left(-\frac{\pi}{2}\right) &\rightsquigarrow \frac{\omega_\varepsilon \sin \mu}{2} \left[(a^2 - b^2) \left(1 - \tilde{\mathcal{B}}_2 \right) + (c+d)(a-b) \left(1 + \tilde{\mathcal{B}}_2 \right) \right]. \end{aligned}$$

Signs of neither “diffusive” nor “mixed” terms are affected by the occurrence of a varying $\mathbf{V}(x, y)$. The divergence-free property rewrites as follows: $\forall (i, j) \in \mathbb{Z}^2$,

$$\begin{aligned} \delta[\omega_\varepsilon \cos \mu]_{i,j} + \delta[\omega_\varepsilon \sin \mu]_{i,j} &:= \left((\omega \cos \mu)_{i+\frac{1}{2}, j+\frac{1}{2}} - (\omega \cos \mu)_{i-\frac{1}{2}, j-\frac{1}{2}} \right) \\ &\quad + \left((\omega \sin \mu)_{i-\frac{1}{2}, j+\frac{1}{2}} - (\omega \sin \mu)_{i+\frac{1}{2}, j-\frac{1}{2}} \right) = 0. \end{aligned}$$

The computation is again split into several steps:

(1) One telescopic sum:

$$\begin{aligned} &\bullet (a+b)(c-d) \left(1 - \tilde{\mathcal{B}}_2 \right) + (c^2 - d^2) \left(1 + \tilde{\mathcal{B}}_2 \right) \\ &\quad = 2(c^2 - d^2) + (c-d) \left[(a+b) - (c+d) \right] (1 - \tilde{\mathcal{B}}_2), \\ &\bullet (c+d)(a-b) \left(1 + \tilde{\mathcal{B}}_2 \right) + (a^2 - b^2) \left(1 - \tilde{\mathcal{B}}_2 \right) \\ &\quad = 2(a^2 - b^2) - (a-b) \left[(a+b) - (c+d) \right] (1 + \tilde{\mathcal{B}}_2). \end{aligned}$$

The first terms add into,

$$(c^2 - d^2) \omega_\varepsilon \cos \mu + (a^2 - b^2) \omega_\varepsilon \sin \mu,$$

a quantity which vanishes thanks to $\mathbf{V}(x, y)$ being divergence-free:

$$\sum_{i,j} |u_{i,j}|^2 \left(\delta[\omega_\varepsilon \cos \mu]_{i,j} + \delta[\omega_\varepsilon \sin \mu]_{i,j} \right) = 0.$$

(2) Second derivative:

$$\begin{aligned} \frac{d}{dt} \left(\sum_{i,j} \frac{|u_{i,j}|^2}{2} \right) &\leq \sum_{i,j} \omega \left[(a+b) - (c+d) \right] \\ &\quad \times \left\{ \frac{c-d}{2R} (1 - \tilde{\mathcal{B}}_2) \cos \mu + \frac{b-a}{2R} (1 + \tilde{\mathcal{B}}_2) \sin \mu \right\}. \end{aligned}$$

The key observation is that $(a+b) - (c+d)$ is a second (cross) derivative:

$$|(a-c) + (b-d)| \leq \Delta x^2 |\partial_{xy}^2 u|.$$

Indeed, by recalling (18), it comes

$$\begin{aligned} a-c &= u_{i-1,j} - u_{i,j} \simeq -\Delta x (\partial_x u)_{i-\frac{1}{2},j}, \\ b-d &= u_{i,j-1} - u_{i-1,j-1} \simeq \Delta x (\partial_x u)_{i-\frac{1}{2},j-1}, \\ (a-c) + (b-d) &\simeq -\Delta x \left((\partial_x u)_{i-\frac{1}{2},j} - (\partial_x u)_{i-\frac{1}{2},j-1} \right) \simeq \Delta x^2 (\partial_{xy}^2 u)_{i-\frac{1}{2},j-\frac{1}{2}}. \end{aligned}$$

(3) Finalized estimate:

$$\left| \frac{c-d}{2R} (1 - \tilde{\mathcal{B}}_2) \omega \cos \mu + \frac{b-a}{2R} (1 + \tilde{\mathcal{B}}_2) \omega \sin \mu \right| \leq \|\mathbf{V}\|_\infty \|\nabla u\|,$$

as $\omega = \|\mathbf{V}\|/2$, so that,

$$\frac{d}{dt} \left(\sum_{i,j} \frac{|u_{i,j}|^2}{2} \right) \leq \Delta x^2 \|\mathbf{V}\|_\infty \sum_{i,j} \left(|\nabla u| |\partial_{xy}^2 u| \right)_{i-\frac{1}{2},j-\frac{1}{2}}.$$

Proposition 4. *Let $\mathbf{V} \in L^\infty(\mathbb{R}^2)$ be divergence-free and let the exact solution be in $H^2(\mathbb{R}^2)$ for $t \geq 0$; then the (semi-discrete) Bessel scheme satisfies*

$$\forall \varepsilon > 0, \quad \frac{d}{dt} \left(\Delta x^2 \sum_{i,j} \frac{|u_{i,j}(t)|^2}{2} \right) \leq \Delta x^2 \|\mathbf{V}\|_\infty \|u(t)\|_{H^2(\mathbb{R}^2)}^2.$$

Remark 7. If the flow is “transport-dominated”, which corresponds to

$$\omega_\varepsilon R \gg 1, \quad \mathcal{B}_2(\omega_\varepsilon R) = 1 + \frac{I_2}{I_0}(\omega_\varepsilon R) \simeq 2,$$

then an interesting factorization can be established as follows:

- Since $\cos 2\mu = 1 - 2\sin^2 \mu = 2\cos^2 \mu - 1$, it comes

$$\cos \mu (1 - \cos 2\mu) = 2\cos \mu \sin^2 \mu, \quad \sin \mu (1 + \cos 2\mu) = 2\sin \mu \cos^2 \mu,$$

so that the error terms rewrite, in this regime,

$$\left[(a+b) - (c+d) \right] \left\{ \frac{c-d}{2R} \omega \cos^2 \mu + \frac{b-a}{2R} \omega \sin^2 \mu \right\} \sin 2\mu;$$

- Yet, by elementary trigonometric relations,

$$\begin{aligned}
 & \frac{\omega \sin 2\mu}{2R} \left\{ (c-d) \cos^2 \mu + (b-a) \omega \sin^2 \mu \right\} \\
 &= \frac{\omega \sin 2\mu}{2R} \left\{ (c-d) + ((b-a) - (c-d)) \sin^2 \mu \right\} \\
 &= \frac{\omega \sin 2\mu}{2R} \left\{ (b-a) - ((b-a) - (c-d)) \cos^2 \mu \right\} \\
 &= \frac{\omega \sin 2\mu}{4R} \left\{ (c-d) + (b-a) - ((b-a) - (c-d)) (\cos^2 \mu - \sin^2 \mu) \right\} \\
 &= -\frac{\omega \sin 2\mu}{4R} ((a-b) - (c-d)) + \frac{\omega \sin 4\mu}{8R} ((a-b) + (c-d)),
 \end{aligned}$$

because

$$(\cos^2 \mu - \sin^2 \mu) \sin 2\mu = \cos 2\mu \sin 2\mu = \frac{\sin 4\mu}{2};$$

which means that there is no enstrophy growth (i.e., no change in the L^2 norm) when $\mu = k\pi/2$, $k \in \mathbb{N}$, because $\sin 2\mu = 0$, but this growth may also be reduced when $\mu = k\pi/4$, as $\sin 4\mu = 0$.

4. CONCLUSION AND OUTLOOK

Both Definition 1 and Theorem 1 allow us to put on a firm ground the notion of “two-dimensional well-balancing” for linear (or weakly nonlinear) equations of the type (1): the exactness at steady-state, that we’re used to in one space dimension, leaves the stage to spectral accuracy for two-dimensional, smooth enough, stationary regimes. Such a property is established for a finite-difference discretization given in [10], which is rather simple to code, and is flexible enough to handle various types of boundary conditions. Yet, this “2D Bessel scheme” is by no means the only one which may turn out to meet with the requirements of Definition 1: in 2D, the first occurrence appears to trace back to Gartland’s “discrete weighted mean” method; see [13] and our Appendix B. Gartland’s ideas were further developed in [3, 37] and later rephrased under the terminology “tailored schemes” in [22]; see also [7]. All these schemes rely on a “four points method of rectangles” approximation of a circular integral involving a Dirichlet-Green function. Such discretizations are closely related to earlier ones, so-called “uniformly accurate” (see [9, 14, 15, 30, 33] and the book [32]). In such finite-difference algorithms, \mathcal{L} -splines (see [35, Chap. 9]) can play an important role, as underlined in [18–20]. Hopefully, spectral error estimates at steady-state will be derived for these numerical methods, too. A greater challenge is to push this analysis toward three-dimensional problems.

APPENDIX A. SPECTRAL ACCURACY OF TRAPEZOIDAL RULE

Let γ_R be the “circular trace operator” $H^s(\mathbb{R}^2) \rightarrow H^{s-\frac{1}{2}}(0, 2\pi R)$ for which

$$\gamma_R : u \mapsto \gamma_R[u], \quad \gamma_R[u](\theta) = u(R \cos \theta, R \sin \theta), \quad f(x) = \gamma_R[u](R\theta),$$

where $\theta \in (0, 2\pi)$. The proof of Lemma 2 is split into several steps:

- Being $f : (0, T = 2\pi R) \rightarrow \mathbb{R}$ periodic, it rewrites as a Fourier series,

$$f(x) = \sum_{k \in \mathbb{Z}} \hat{f}(k) \exp(ik \frac{2\pi x}{T}), \quad \hat{f}(k) = \frac{1}{T} \int_0^T f(x) \exp(-ik \frac{2\pi x}{T}) dx.$$

Accordingly, the N -points (Nyström's) "method of rectangles" rewrites:

$$\begin{aligned} \frac{T}{N} \sum_{j=0}^{N-1} f\left(j \frac{T}{N}\right) &= \frac{T}{N} \sum_{j=0}^{N-1} \left(\sum_{k \in \mathbb{Z}} \hat{f}(k) \exp\left(ik \frac{2\pi j}{N}\right) \right) \\ &= \frac{T}{N} \sum_{k \in \mathbb{Z}} \hat{f}(k) \left(\sum_{j=0}^{N-1} \exp\left(ik \frac{2\pi j}{N}\right) \right). \end{aligned}$$

- An elementary observation is that

$$\frac{1}{T} \int_0^T \exp\left(ik \frac{2\pi x}{T}\right) dx - \frac{1}{N} \sum_{j=0}^{N-1} \exp\left(ik \frac{2\pi j}{N}\right) = -\chi(k \in N\mathbb{Z}_*),$$

where χ stands for the indicator function and \mathbb{Z}_* , for \mathbb{Z} without 0. Indeed, for $k \in \mathbb{Z}$, being a geometric summation of reason $\exp\left(ik \frac{2\pi}{N}\right)$,

$$\sum_{j=0}^{N-1} \exp\left(ik \frac{2\pi j}{N}\right) = \frac{1 - \exp\left(ik \frac{2\pi}{N}\right)}{1 - \exp\left(ik \frac{2\pi}{N}\right)} = N \chi(k \in N\mathbb{Z}),$$

so the numerator always vanishes. Indexes $k \in N\mathbb{Z}$ are such that the denominator cancels, too: a Taylor expansion yields the expected result. Yet,

$$(20) \quad \hat{f}(0) - \left(\frac{1}{N} \sum_{k \in \mathbb{Z}} \hat{f}(k) \sum_{j=0}^{N-1} \exp\left(ik \frac{2\pi j}{N}\right) \right) = - \sum_{k \in \mathbb{Z}, k \neq 0} \hat{f}(Nk),$$

and it remains to estimate the lacunary sum on the right-hand side.

- To estimate the right-hand side of (20), we recall that f is the restriction on a circle of radius $R > 0$ and centered in (x_0, y_0) of $u(x, y) \in H^s(\mathbb{R}^2)$. Standard trace theory [29] gives that $f \in H^{s-\frac{1}{2}}(0, T)$: if $s > 1$, Sobolev inequalities ensure that f is continuous. Moreover, for any index $k \in \mathbb{N}$,

$$\begin{aligned} \hat{f}(k) &= \frac{1}{2\pi} \int_0^{2\pi} u(R \cos \theta, R \sin \theta) \exp(-ik\theta) d\theta \\ &= -\frac{R}{2\pi} \int_0^{2\pi} \left(-\sin \theta \partial_x u + \cos \theta \partial_y u \right) \left(\frac{-1}{ik} \exp(-ik\theta) \right) d\theta \\ &= \frac{-iR}{2\pi k} \int_0^{2\pi} -\partial_x u \left(\sin \theta \exp(-ik\theta) \right) + \partial_y u \left(\cos \theta \exp(-ik\theta) \right) d\theta \\ &= \frac{R}{4\pi k} \int_0^{2\pi} \partial_x u \left(\exp(-i(k-1)\theta) - \exp(-i(k+1)\theta) \right) d\theta \\ &\quad - \frac{iR}{4\pi k} \int_0^{2\pi} \partial_y u \left(\exp(-i(k-1)\theta) + \exp(-i(k+1)\theta) \right) d\theta. \end{aligned}$$

Recalling that $\gamma_R[u](\theta) = u(R \cos \theta, R \sin \theta)$, this yields:

$$\begin{aligned} \widehat{\gamma_R[u]}(k) &= \frac{R}{2k} \left(\widehat{\gamma_R[\partial_x u]}(k-1) - \widehat{\gamma_R[\partial_x u]}(k+1) \right. \\ &\quad \left. - i \left(\widehat{\gamma_R[\partial_y u]}(k-1) + \widehat{\gamma_R[\partial_y u]}(k+1) \right) \right), \end{aligned}$$

so that, the smoothness of u induces Fourier coefficients of $\gamma_R[u]$ which are of the order of powers of R , the radius of the circle. By induction, this calculation can be repeated until

- either Fourier coefficients of index zero are reached, and integrating by parts isn't possible;
- or the function u hasn't enough regularity (when $|k| > s - \frac{1}{2}$).

Finally, we get an estimate on the size of Fourier coefficients on the circle of radius R , which can obviously be used to control the error (20):

$$(21) \quad \boxed{\forall k \in \mathbb{Z}, \quad \left| \widehat{\gamma_R[u]}(k) \right| = O \left(R^{\min(|k|, s - \frac{1}{2})} \right).}$$

If $s - \frac{1}{2} \geq |k|$, then in general,

$$\forall k \in \mathbb{Z}, \quad \left| \widehat{\gamma_R[u]}(k) \right| = O(R^{|k|}) \left| \widehat{\gamma_R[D^{|k|}u]}(0) \right|,$$

which is the “circular average” of $D^{|k|}u$, and the value at the circle's center of its harmonic extension. This agrees with the approximate values found in [10, eq. (3.7)] used in (11). Related estimates also appear in [36] for different purposes.

APPENDIX B. ACCURACY OF “DISCRETE WEIGHTED MEANS”

Lemma 2 allows us to prove spectral accuracy for the so-called “discrete weighted mean approximations” of strictly elliptic equations as well (see [13], later [3, 37], and also the “tailored schemes” [22]). In its simplest occurrence, it deals with a two-dimensional “modified Helmholtz” (or “steady Klein-Gordon”) equation,

$$(22) \quad -\varepsilon^2 \Delta u + \alpha^2 u = 0, \text{ in } \Omega, \quad u|_{\partial\Omega} = \varphi, \quad \varepsilon > 0,$$

with $\alpha \in \mathbb{R}_*^+$ and Ω being a bounded polygonal domain in \mathbb{R}^2 , for instance the unit square. It consists in building U , a finite-difference approximation of (22) by

- (1) covering the computational domain with disks $D_{\Delta x}(i, j)$; see [37, Fig. 2];
- (2) approximating each value at (x_i, y_j) by applying the “4-points method of rectangles” to the integral formulation in each of the disks $D_{\Delta x}(i, j)$ by means of the Dirichlet-Green function, $H_\varepsilon^0(r, \theta; \rho, \psi)$.

From [10, §2.4], the Dirichlet-Green function for (22) in $D_{\Delta x}(i, j)$ reads:

$$H_\varepsilon^0(r = 0, \theta; \rho = \Delta x, \psi) = \frac{1}{2\pi \Delta x I_0(\alpha_\varepsilon \Delta x)}, \quad \alpha_\varepsilon = \frac{\alpha}{\varepsilon},$$

so that we have the (exact) integral representation: $\forall(i, j)$,

$$\begin{aligned} u(x_i, y_j) &= \frac{1}{2\pi \Delta x I_0(\alpha_\varepsilon \Delta x)} \int_{C_{\Delta x}} u(x, y) \, d\vec{n} \\ &= \frac{1}{2\pi I_0(\alpha_\varepsilon \Delta x)} \int_0^{2\pi} u(\Delta x \cos \psi, \Delta x \sin \psi) \, d\psi \\ &\simeq \frac{1}{4 I_0(\alpha_\varepsilon \Delta x)} \left(u(x_i + \Delta x, y_j) + u(x_i, y_j + \Delta x) \right. \\ &\quad \left. + u(x_i - \Delta x, y_j) + u(x_i, y_j - \Delta x) \right), \end{aligned}$$

the last line being obtained by approximating the integral with the four discrete values that we have at hand. Clearly, this defines a numerical scheme for (22), which is the one written in [37, Theorem 2: eq. (2)].

Proposition 5. Let $u \in H^{\frac{9}{2}}(\mathbb{R}^2)$ be the solution of $-\varepsilon \Delta u + \alpha^2 u = 0$ in $D_{\Delta x}(i, j)$; then

$$\left| u(x_i, y_j) - \sum_{k=0}^3 \frac{u(x_i + \Delta x \cos(k\frac{\pi}{2}), y_j + \Delta x \sin(k\frac{\pi}{2}))}{4 I_0(\alpha_\varepsilon \Delta x)} \right| \leq \frac{O(\Delta x^4)}{I_0(\alpha_\varepsilon \Delta x)}.$$

Real coefficients aren't necessary for the above derivation of "discrete weighted means". Picking a purely imaginary α , (22) becomes a Helmholtz equation, and $I_0(\alpha_\varepsilon \Delta x) = J_0(|\alpha_\varepsilon| \Delta x)$: an "optimal 2D scheme" studied in [39, eq. (25)] is found.

Proof. Let $T = 2\pi \Delta x$ be the perimeter of $D_{\Delta x}(i, j)$. The exact expression of $u(x_i, y_j)$, based on data on the circle $C_{\Delta x}(i, j)$, rewrites (using polar coordinates)

$$\begin{aligned} u(x_i, y_j) &= \frac{1}{2\pi \Delta x} \int_0^{2\pi \Delta x} \frac{\tilde{u}(\Delta x, \psi)}{I_0(\alpha_\varepsilon \Delta x)} d\psi. \quad \text{where } \tilde{u}(r, \psi) = u(r \cos \psi, r \sin \psi), \\ &= \frac{1}{2\pi I_0(\alpha_\varepsilon \Delta x)} \int_0^{2\pi} \tilde{u}(\Delta x, \psi) d\psi \\ &= \frac{1}{I_0(\alpha_\varepsilon \Delta x)} \left[\frac{1}{4} \sum_{k=0}^3 \tilde{u}(\Delta x, k\frac{\pi}{2}) + O(\Delta x^4) \right], \end{aligned}$$

because the trace on the circle $C_{\Delta x}(i, j)$ belongs to $H^{\frac{9}{2}-\frac{1}{2}}(0, T) = H^4(0, T)$. \square

It agrees with an estimate, based on Taylor expansions [37, Theorem 8],

$$\left| u(x_i, y_j) - \sum_{k=0}^3 \frac{u(x_i + \Delta x \cos(k\frac{\pi}{2}), y_j + \Delta x \sin(k\frac{\pi}{2}))}{4 I_0(\alpha_\varepsilon \Delta x)} \right| \leq \frac{O(\Delta x^4)}{I_0(\alpha_\varepsilon \Delta x)}.$$

Accordingly, for $u \in H^{\frac{9}{2}}(\Omega)$ solution of (22), the global L^2 bound stated in [37, Theorem 2; eq. (3)] improves into (with the same notation as the authors),

$$(23) \quad \boxed{\|\hat{u} - U\|_{L^2(\Omega)} := \Delta x^2 \|\hat{u} - U\|_{\ell^2} \leq \frac{O(\Delta x^4)}{I_0(\alpha_\varepsilon \Delta x) - \cos(\pi \Delta x)},}$$

being \hat{u} the restriction of the exact solution u of (22) to discrete grid points (x_i, y_j) .

Remark 8. Since $H_\varepsilon^0(0, \theta; \Delta x, \psi)$ is independent on ψ , the standard finite-difference scheme for (22) simply consists in substituting, inside the "four-points method of rectangles", the modified Bessel function $I_0(\alpha_\varepsilon \Delta x)$ by its polynomial approximation,

$$I_0(\alpha_\varepsilon \Delta x) \simeq T_0(\alpha_\varepsilon \Delta x) := 1 + \frac{|\alpha_\varepsilon \Delta x|^2}{4}, \quad \alpha_\varepsilon \Delta x \ll 1.$$

Accordingly, a bound as strong as the one stated in Proposition 5 cannot hold for finite-differences because its local truncation error [11] contains a supplementary term,

$$(24) \quad \|u\|_\infty \left| \frac{1}{I_0(\alpha_\varepsilon \Delta x)} - \frac{1}{T_0(\alpha_\varepsilon \Delta x)} \right| \leq \frac{O(|\alpha_\varepsilon \Delta x|^4) \|u\|_\infty}{I_0(\alpha_\varepsilon \Delta x) T_0(\alpha_\varepsilon \Delta x)} \quad \alpha_\varepsilon \Delta x \rightarrow 0.$$

If $\alpha_\varepsilon \Delta x \gg 1$, (23) produces a smaller bound than (24), hence a better accuracy.

In the unit square $\Omega = (0, 1) \times (0, 1)$ it is easy to find an exact solution of (22),

$$(25) \quad u(x, y) = \cosh(\alpha_\varepsilon x) + \sinh(\alpha_\varepsilon y),$$

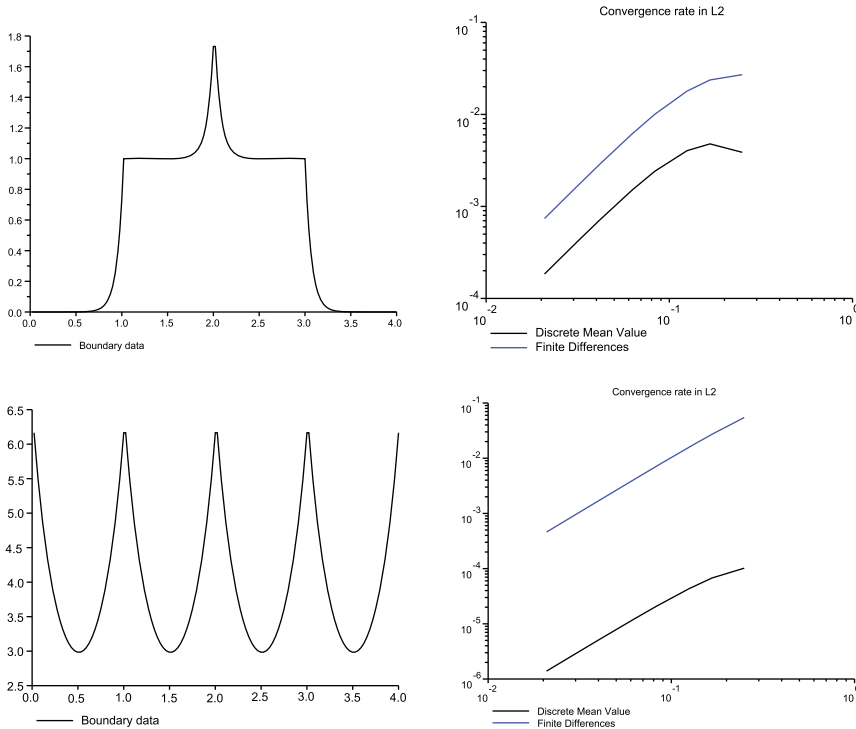


FIGURE 5. Boundary data (left) and measured L^2 convergence rates (right) with exact solutions (25) and (26). (top to bottom)

so that a numerical test can be set up by confronting finite-differences and “discrete weighted means” with $\alpha_\varepsilon^2 = 225 = 15^2$. Another is a radial modified Bessel function,

$$(26) \quad u(x, y) = I_0 \left(\alpha_\varepsilon \sqrt{\left(x - \frac{1}{2}\right)^2 + \left(y - \frac{1}{2}\right)^2} \right),$$

These exact solutions induce nonsmooth boundary data, continuous on $\partial\Omega$, but not differentiable at corners, so that “corner singularities” [23] are expected because “compatibility conditions” aren’t met [4]. Convergence rate is measured on Figure 5: second-order accuracy holds in L^2 , in agreement with [37, §6] and [39, §3.2]. Moreover, global errors generated by DWM are much lower with (26) than (25).

A more elaborate example is the one originally proposed by Gartland; see [13, §3]: it considers the two-dimensional advection-diffusion equation,

$$(27) \quad -\varepsilon \Delta u + \partial_x u = 0, \quad \text{in } \Omega, \quad u|_{\partial\Omega} = \varphi,$$

to which exactly the same procedure is applied. In agreement with [10, §2.4], in any disk $D_{\Delta x}(i, j)$ the exact solution at the center reads, using again polar coordinates,

$$u(x_i, y_j) = \int_0^{2\pi} \tilde{u}(\Delta x, \psi) \frac{\exp\left(-\frac{\Delta x}{2\varepsilon} \cos \psi\right)}{2\pi I_0(\Delta x/2\varepsilon)} d\psi,$$

which is an integral to which the “4-points method of rectangles” can be applied; see [13, eq. (3.3)], so that Lemma 2 yields local spectral accuracy. Accordingly, a global L^2 estimate, somehow similar to (23), may be expected to hold for (27), too.

APPENDIX C. ELEMENTARY TRACE ESTIMATES ON A CIRCLE

Let $R > 0$ and D stand for the (open) disk centered in $\vec{0}$ and radius $R > 0$; C is the circle of identical radius, so that $C = \partial D$.

C.1. L^2 estimate. Let $f \in C^1$ and denote $\ell = R\theta$ as its the arc-length,

$$\frac{1}{R} \frac{dg}{d\theta}(\theta) = \frac{dg}{d\ell}(\ell) = (-\sin \theta, \cos \theta) \cdot \nabla_{x,y} f(x, y).$$

Polar coordinates $x = r \cos \theta$, $y = r \sin \theta$, $0 \leq r \leq R$, yield:

$$\nabla_{x,y} f(x, y) = (\cos \theta, \sin \theta) \frac{\partial f}{\partial r} + \frac{1}{r} (-\sin \theta, \cos \theta) \frac{\partial f}{\partial \theta},$$

so that the H^1 norm in D reads

$$\begin{aligned} \|f\|_{H^1(D)}^2 &= \int_D f(x, y)^2 + |\nabla_{x,y} f|^2 \, dx \, dy \\ &= \int_0^R \int_0^{2\pi} f(r, \theta)^2 + \left| \frac{\partial f}{\partial r} \right|^2 + \left| \frac{1}{r} \frac{\partial f}{\partial \theta} \right|^2 r \, dr \, d\theta, \end{aligned}$$

along with the L^2 norm on C :

$$\|g\|_{L^2(C)}^2 = \int_0^{2\pi R} g(\ell)^2 \, d\ell = \int_0^{2\pi} f(R, \theta)^2 R \, d\theta.$$

This 1D integral on C is meant to be bounded by means of $2D$ integrals involving the whole interval $r \in (0, R)$. Hence,

$$\begin{aligned} \forall R > 0, \quad R^2 |f(R, \theta)|^2 &= \int_0^R \frac{\partial}{\partial r} (r^2 f^2(r, \theta)) \, dr \\ &= 2 \int_0^R \left(r f \frac{\partial f}{\partial r} + f^2 \right) r \, dr. \end{aligned}$$

By integrating on $\theta \in (0, 2\pi)$, and using the Cauchy-Schwarz inequality twice,

$$\begin{aligned} \frac{R}{2} \int_0^{2\pi} |f(R, \theta)|^2 R \, d\theta &\leq \|r f\|_{L^2(D_R)} \left\| \frac{\partial f}{\partial r} \right\|_{L^2(D_R)} + \|f\|_{L^2(D_R)}^2 \\ &\leq \sqrt{\frac{\pi}{2}} R^2 \|f\|_{L^2(D_R)} \left\| \frac{\partial f}{\partial r} \right\|_{L^2(D_R)} + \|f\|_{L^2(D_R)}^2. \end{aligned}$$

A sharp (equality for constants; see [6]) trace estimate is

$$(28) \quad \boxed{\|g\|_{L^2(C)}^2 \leq \frac{2}{R} \|f\|_{L^2(D)}^2 + R\sqrt{2\pi} \|f\|_{L^2(D)} \left\| \frac{\partial f}{\partial r} \right\|_{L^2(D)}}.$$

Being any function of $H^1(D)$ also in $L^q(D_R)$ for $2 \leq q < +\infty$,

$$\begin{aligned} \frac{1}{R^2} \int_D f^2(x, y) \, dx \, dy &= \frac{1}{R^2} \int_{\mathbb{R}^2} \chi_D(\chi_D f^2) \, dx \, dy \\ &\leq \frac{1}{R^2} (\pi R^2)^{\frac{1}{p}} \|\chi_D f^2\|_{L^q} \quad \left(\text{H\"older, } \frac{1}{p} + \frac{1}{q} = 1\right) \\ &\leq \frac{1}{R^2} (\pi R^2)^{\frac{1}{p}} \|f\|_{L^{2q}(D)}^2 \\ &\leq O(R^{\frac{2}{p}-2}) \|f\|_{L^{\frac{2p}{p-1}}(D)}^2 \\ &\leq O(R^{-\frac{2}{q}}) \|f\|_{L^{2q}(D)}^2, \end{aligned}$$

(where χ_D is the indicator function of the disk D) so that (28) rewrites:

$$\frac{\|g\|_{L^2(C)}^2}{R} \leq O(R^{-\frac{2}{q}}) \|f\|_{L^{2q}(D_R)}^2 + \sqrt{2\pi} \|f\|_{L^2(D)} \left\| \frac{\partial f}{\partial r} \right\|_{L^2(D)}, \quad q \geq 1.$$

It remains to invoke a standard density argument.

C.2. $H^{\frac{1}{2}}$ estimate. If $f \in H^1(\bar{D})$, then g , its trace on the circle C of radius R is well-defined as an $H^{\frac{1}{2}}$ function [29, Lemma 4], for which a scaling argument yields:

$$|g|_{H^{\frac{1}{2}}(C)} \leq C \left(\frac{1}{R} \|f\|_{L^2(D)} + \|\nabla f\|_{L^2(D)} \right).$$

The rescaling proceeds by introducing F defined on the unit disk,

$$f(x, y) = F(R\tilde{x}, R\tilde{y}), \quad |\tilde{x}|^2 + |\tilde{y}|^2 \leq 1,$$

and computing that, in $n = 2$ space dimensions,

- $\|f\|_p = R^{\frac{n}{p}} \|F\|_p$, thus $\|f\|_{L^2(D)} = R \|F\|_{L^2}$,
- $\|\nabla f\|_p = R^{\frac{n}{p}-1} \|\nabla F\|_p$, thus $\|\nabla f\|_{L^2(D)} = \|\nabla F\|_{L^2}$,
- $\|g\|_p = R^{(n-1)/p} \|G\|_p$, thus $\|g\|_{L^2(C)} = R^{\frac{1}{2}} \|G\|_{L^2}$,
- $|g|_{B_{p,p}^{1-\frac{1}{p}}} = R^{\frac{n}{p}-1} |G|_{B_{p,p}^{1-\frac{1}{p}}}$ for Besov semi-norms of g and G , which are the traces of f and F , so that $|g|_{H^{\frac{1}{2}}(C)} = |g|_{H^{\frac{1}{2}}}$.

The estimates given in [29, Lemma 4], which concern F and G ,

$$\begin{aligned} \|G\|_p &\leq C(\|F\|_p + \|\nabla F\|_p), \\ |G|_{B_{p,p}^{1-\frac{1}{p}}} &\leq C(\|F\|_p + \|\nabla F\|_p), \end{aligned}$$

rewrite as estimates acting on f and g by rescaling,

$$\begin{aligned} R^{-\frac{1}{2}} \|g\|_{L^2(C)} &\leq C(R^{-1} \|f\|_{L^2(D)} + \|\nabla f\|_{L^2(D)}), \\ |g|_{H^{\frac{1}{2}}(C)} &\leq C(R^{-1} \|f\|_{L^2(D)} + \|\nabla f\|_{L^2(D)}). \end{aligned}$$

Again, any function of $H^1(D)$ is also in $L^q(D)$ for $2 \leq q < +\infty$, so

$$\begin{aligned} \frac{1}{R^2} \int_D f^2(x, y) \, dx \, dy &= \frac{1}{R^2} \int_{\mathbb{R}^2} \chi_D (\chi_D f^2) \, dx \, dy \\ &\leq \frac{1}{R^2} (\pi R^2)^{\frac{1}{p}} \|\chi_D f^2\|_{L^q} \quad \left(\text{Hölder, } \frac{1}{p} + \frac{1}{q} = 1\right) \\ &\leq \frac{1}{R^2} (\pi R^2)^{\frac{1}{p}} \|f\|_{L^{2q}(D)}^2 \\ &\leq O(R^{\frac{2}{p}-2}) \|f\|_{L^{\frac{2p}{p-1}}(D)}^2 \\ &\leq O(R^{-\frac{2}{q}}) \|f\|_{L^{2q}(D)}^2. \end{aligned}$$

Finally, the (semi-norm) trace estimate reduces to

$$(29) \quad |g|_{H^{\frac{1}{2}}(C)} \leq C \left(\frac{1}{R^{\frac{1}{q}}} \|f\|_{L^{2q}(D)} + \|\nabla f\|_{L^2(D)} \right), \quad q \geq 1,$$

so the effect of the singular term $\frac{1}{R}$ in the general estimate weakens with $q \gg 1$.

ACKNOWLEDGMENTS

The author thanks Dr. Andrea Di Mascio for numerical coding in FORTRAN, along with Professors François Bouchut and Yoichi Miyazaki.

REFERENCES

- [1] M. Ainsworth and W. Dörfler, *Fundamental systems of numerical schemes for linear convection-diffusion equations and their relationship to accuracy*, Computing **66** (2001), no. 2, 199–229, DOI 10.1007/s006070170035. Archives for scientific computing. Numerical methods for transport-dominated and related problems (Magdeburg, 1999). MR1825804
- [2] D. Amadori and L. Gosse, *Error estimates for well-balanced schemes on simple balance laws*, SpringerBriefs in Mathematics, Springer, Cham; BCAM Basque Center for Applied Mathematics, Bilbao, 2015. One-dimensional position-dependent models; With a foreword by François Bouchut; BCAM SpringerBriefs. MR3442979
- [3] M. G. Andrade and J. B. R. do Val, *A numerical scheme based on mean value solutions for the Helmholtz equation on triangular grids*, Math. Comp. **66** (1997), no. 218, 477–493, DOI 10.1090/S0025-5718-97-00825-9. MR1401937
- [4] V. B. Andreev, *On the accuracy of grid approximations of nonsmooth solutions of a singularly perturbed reaction-diffusion equation in the square* (Russian, with Russian summary), Differ. Uravn. **42** (2006), no. 7, 895–906, 1005, DOI 10.1134/S0012266106070044; English transl., Differ. Equ. **42** (2006), no. 7, 954–966. MR2294140
- [5] A. Arakawa, *Computational design for long-term numerical integration of the equations of fluid motion: two-dimensional incompressible flow. I* [J. Comput. Phys. **1** (1966), no. 1, 119–143], J. Comput. Phys. **135** (1997), no. 2, 101–114, DOI 10.1006/jcph.1997.5697. With an introduction by Douglas K. Lilly; Commemoration of the 30th anniversary [of J. Comput. Phys.]. MR1486265
- [6] G. Auchmuty, *Sharp boundary trace inequalities*, Proc. Roy. Soc. Edinburgh Sect. A **144** (2014), no. 1, 1–12, DOI 10.1017/S0308210512000601. MR3164533
- [7] O. Axelsson, E. Glushkov, and N. Glushkova, *The local Green's function method in singularly perturbed convection-diffusion problems*, Math. Comp. **78** (2009), no. 265, 153–170, DOI 10.1090/S0025-5718-08-02161-3. MR2448701
- [8] O. Axelsson and W. Layton, *Defect correction methods for convection-dominated convection-diffusion problems* (English, with French summary), RAIRO Modél. Math. Anal. Numér. **24** (1990), no. 4, 423–455, DOI 10.1051/m2an/1990240404231. MR1070965

- [9] A. E. Berger, J. M. Solomon, and M. Ciment, *An analysis of a uniformly accurate difference method for a singular perturbation problem*, Math. Comp. **37** (1981), no. 155, 79–94, DOI 10.2307/2007501. MR616361
- [10] R. Bianchini and L. Gosse, *A truly two-dimensional discretization of drift-diffusion equations on Cartesian grids*, SIAM J. Numer. Anal. **56** (2018), no. 5, 2845–2870, DOI 10.1137/17M1151353. MR3855393
- [11] G. Birkhoff and S. Gulati, *Optimal few-point discretizations of linear source problems*, SIAM J. Numer. Anal. **11** (1974), 700–728, DOI 10.1137/0711057. MR0362933
- [12] C. L. Epstein, *How well does the finite Fourier transform approximate the Fourier transform?*, Comm. Pure Appl. Math. **58** (2005), no. 10, 1421–1435, DOI 10.1002/cpa.20064. MR2162785
- [13] E. C. Gartland Jr., *Discrete weighted mean approximation of a model convection-diffusion equation*, SIAM J. Sci. Statist. Comput. **3** (1982), no. 4, 460–472, DOI 10.1137/0903030. MR677099
- [14] E. C. Gartland Jr., *Strong stability of compact discrete boundary value problems via exact discretizations*, SIAM J. Numer. Anal. **25** (1988), no. 1, 111–123, DOI 10.1137/0725009. MR923929
- [15] E. C. Gartland Jr., *On the uniform convergence of the Scharfetter-Gummel discretization in one dimension*, SIAM J. Numer. Anal. **30** (1993), no. 3, 749–758, DOI 10.1137/0730037. MR1220650
- [16] U. Ghia, K.N. Ghia, C.T. Shin, *High-Re solutions for incompressible flow using the Navier-Stokes equations and a multigrid method*, J. Comput. Phys. **48** (1982) 387–411.
- [17] L. Gosse, *Computing Qualitatively Correct Approximations of Balance Laws*, SIMAI Springer Series, vol. 2, Springer, Milan, 2013. Exponential-fit, well-balanced and asymptotic-preserving. MR3053000
- [18] L. Gosse, *Dirichlet-to-Neumann mappings and finite-differences for anisotropic diffusion*, Comput. & Fluids **156** (2017), 58–65, DOI 10.1016/j.compfluid.2017.06.026. MR3693203
- [19] L. Gosse, *Viscous equations treated with \mathcal{L} -splines and Steklov-Poincaré operator in two dimensions*, in Innovative Algorithms & Analysis, DOI: 10.1007/978-3-319-49262-9_6.
- [20] L. Gosse, *\mathcal{L} -splines and viscosity limits for well-balanced schemes acting on linear parabolic equations*, Acta Appl. Math. **153** (2018), 101–124, DOI 10.1007/s10440-017-0122-5. MR3745732
- [21] J. M. Greenberg and A. Y. Leroux, *A well-balanced scheme for the numerical processing of source terms in hyperbolic equations*, SIAM J. Numer. Anal. **33** (1996), no. 1, 1–16, DOI 10.1137/0733001. MR1377240
- [22] H. Han, Z. Huang, and R. B. Kellogg, *A tailored finite point method for a singular perturbation problem on an unbounded domain*, J. Sci. Comput. **36** (2008), no. 2, 243–261, DOI 10.1007/s10915-008-9187-7. MR2434846
- [23] H. Han and R. B. Kellogg, *Differentiability properties of solutions of the equation $-\epsilon^2 \Delta u + ru = f(x, y)$ in a square*, SIAM J. Math. Anal. **21** (1990), no. 2, 394–408, DOI 10.1137/0521022. MR1038899
- [24] J. P. Hennart, *On the numerical analysis of analytical nodal methods*, Numer. Methods Partial Differential Equations **4** (1988), no. 3, 233–254, DOI 10.1002/num.1690040306. MR1012482
- [25] W. Layton, *Optimal difference schemes for 2-D transport problems*, J. Comput. Appl. Math. **45** (1993), no. 3, 337–341, DOI 10.1016/0377-0427(93)90051-C. MR1216076
- [26] K. Lipnikov, G. Manzini, and M. Shashkov, *Mimetic finite difference method*. part B, J. Comput. Phys. **257** (2014), no. part B, 1163–1227, DOI 10.1016/j.jcp.2013.07.031. MR3133437
- [27] P. L. Butzer, G. Schmeisser, and R. L. Stens, *An introduction to sampling analysis*, Nonuniform sampling, Inf. Technol. Transm. Process. Storage, Kluwer/Plenum, New York, 2001, pp. 17–121. MR1875676
- [28] A. Mayo, *Fast high order accurate solution of Laplace’s equation on irregular regions*, SIAM J. Sci. Statist. Comput. **6** (1985), no. 1, 144–157, DOI 10.1137/0906012. MR773287
- [29] Y. Miyazaki, *Sobolev trace theorem and the Dirichlet problem in the unit disk*, Milan J. Math. **82** (2014), no. 2, 297–312, DOI 10.1007/s00032-014-0222-x. MR3277700
- [30] E. O’Riordan and M. Stynes, *A globally uniformly convergent finite element method for a singularly perturbed elliptic problem in two dimensions*, Math. Comp. **57** (1991), no. 195, 47–62, DOI 10.2307/2938662. MR1079029

- [31] Q. I. Rahman and G. Schmeisser, *Characterization of the speed of convergence of the trapezoidal rule*, Numer. Math. **57** (1990), no. 2, 123–138, DOI 10.1007/BF01386402. MR1048307
- [32] H.-G. Roos, M. Stynes, and L. Tobiska, *Robust Numerical Methods for Singularly Perturbed Differential Equations*, 2nd ed., Springer Series in Computational Mathematics, vol. 24, Springer-Verlag, Berlin, 2008. Convection-diffusion-reaction and flow problems. MR2454024
- [33] R. Sacco and M. Stynes, *Finite element methods for convection-diffusion problems using exponential splines on triangles*, Comput. Math. Appl. **35** (1998), no. 3, 35–45, DOI 10.1016/S0898-1221(97)00277-0. MR1605547
- [34] H. L. Scharfetter, H. K. Gummel, *Large signal analysis of a silicon Read diode oscillator*, IEEE Trans. Electron Devices **16** (1969) 64–77.
- [35] L. L. Schumaker, *Spline functions: basic theory*, 3rd ed., Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2007. MR2348176
- [36] E. Tadmor, *The exponential accuracy of Fourier and Chebyshev differencing methods*, SIAM J. Numer. Anal. **23** (1986), no. 1, 1–10, DOI 10.1137/0723001. MR821902
- [37] J. B. R. do Val and M. G. Andrade, *On mean value solutions for the Helmholtz equation on square grids*, Appl. Numer. Math. **41** (2002), no. 4, 459–479, DOI 10.1016/S0168-9274(01)00127-1. MR1905218
- [38] L. N. Trefethen and J. A. C. Weideman, *The exponentially convergent trapezoidal rule*, SIAM Rev. **56** (2014), no. 3, 385–458, DOI 10.1137/130932132. MR3245858
- [39] Y. S. Wong and G. Li, *Exact finite difference schemes for solving Helmholtz equation at any wavenumber*, Int. J. Numer. Anal. Model. Ser. B **2** (2011), no. 1, 91–108. MR2866996
- [40] A. A. Žensykbayev, *Best quadrature formula for the class $W^r L_2$* (English, with Russian summary), Anal. Math. **3** (1977), no. 1, 83–93, DOI 10.1007/BF02333255. MR0447924

IAC–CNR “MAURO PICONE” (SEZIONE DI ROMA) - VIA DEI TAURINI, 19 - 00185 ROME, ITALY
 Email address: l.gosse@ba.iac.cnr.it