SIAM J. NUMER. ANAL.
Vol. 58, No. 2, pp. 1092–1116

# A TRULY TWO-DIMENSIONAL, ASYMPTOTIC-PRESERVING SCHEME FOR A DISCRETE MODEL OF RADIATIVE TRANSFER[*]

LAURENT GOSSE[†] AND NICOLAS VAUCHELET[‡]

**Abstract.** For a four-stream approximation of the kinetic model of radiative transfer with isotropic scattering, a numerical scheme endowed with both truly 2D well-balanced and diffusive asymptotic-preserving properties is derived, in the same spirit as what was done in [L. Gosse and G. Toscani, *C. R. Math. Acad. Sci. Paris*, 334 (2002), pp. 337–342] in the 1D case. Building on former results of Birkhoff and Abu-Shumays [*J. Math. Anal. Appl.*, 28 (1969), pp. 211–221], it is possible to express 2D kinetic steady-states by means of harmonic polynomials, and this allows one to build a scattering $S$-matrix yielding a time-marching scheme. Such an $S$-matrix can be decomposed, as in [L. Gosse and N. Vauchelet, *Numer. Math.*, 141 (2019), pp. 627–680], so as to deduce another scheme, well-suited for a diffusive approximation of the kinetic model, for which rigorous convergence can be proved. Challenging benchmarks are also displayed on coarse grids.

**Key words.** asymptotic-preserving, diffusive scaling, four-stream approximation, grey radiative transfer, $S$-matrix

**AMS subject classifications.** 31A05, 65M06, 76R50, 82B40, 85A25

**DOI.** 10.1137/19M1239829

## 1. Introduction and preliminaries.

### 1.1. Kinetic modeling in 2D.

We are interested in a "truly two-dimensional" numerical simulation of the simple kinetic model, where $\mathbf{x} = (x, y)$ and $\mathbf{v} = (\xi, \eta)$,

$$\partial_t f(t, \mathbf{x}, \mathbf{v}) + \mathbf{v} \cdot \nabla_{\mathbf{x}} f = \sigma(\mathbf{x}) \left( \int_{\mathbb{S}^1} f(t, \mathbf{x}, \mathbf{v}') \frac{\mathrm{d}\mathbf{v}'}{2\pi} - f \right), \qquad |\mathbf{v}| = 1,$$

in particular, of its "four-stream approximation" evoked in, e.g., [19, sect. 5] and [11], where the "opacity" $\sigma(x, y) \geq 0$ and the macroscopic density simplifies into

$$\forall t, \mathbf{x} \in \mathbb{R}^+ \times \mathbb{R}^2, \qquad \rho(t, \mathbf{x}) = f^+(t, \mathbf{x}) + f^-(t, \mathbf{x}) + g^+(t, \mathbf{x}) + g^-(t, \mathbf{x}).$$

In order to take full advantage of a 9-point, so-called *Moore* stencil, microscopic velocities are rotated so as to be aligned with the diagonals of a Cartesian grid,

$$\mathbf{v} \in \left\{ \frac{\pm 1}{\sqrt{2}}(1, 1), \frac{\pm 1}{\sqrt{2}}(-1, 1) \right\}, \tag{1.1}$$

such as, for instance, in [5, sect. 2.1]. This choice leads to the following 2D system:

$$\begin{cases} \partial_t f^\pm \pm \dfrac{1}{\sqrt{2}} \left( \partial_x f^\pm + \partial_y f^\pm \right) = \sigma(x, y) \left( \dfrac{\rho}{4} - f^\pm \right), \\ \partial_t g^\pm \mp \dfrac{1}{\sqrt{2}} \left( \partial_x g^\pm - \partial_y g^\pm \right) = \sigma(x, y) \left( \dfrac{\rho}{4} - g^\pm \right), \end{cases} \tag{1.2}$$

[*] Received by the editors January 24, 2019; accepted for publication (in revised form) February 6, 2020; published electronically April 7, 2020.
https://doi.org/10.1137/19M1239829

**Funding:** The work of the authors was supported by the Italian Minister of Instruction, University of Research (MIUR) with PRIN Project 2017 "Innovative numerical methods for evolutionary partial differential equations and applications" grant 2017KKJP4X.

[†] IAC–CNR "Mauro Picone," Via dei Taurini 19, 00185 Rome, Italy (l.gosse@ba.iac.cnr.it).

[‡] Laboratoire Analyse, Géométrie et Applications (LAGA), Université Sorbonne Paris Nord, CNRS UMR 7539, F-93430, Villetaneuse, France (vauchelet@math.univ-paris13.fr).

for which we propose a numerical scheme endowed with similar properties as the one in [16], in a two-dimensional context, without domain decomposition, like in [2, 18, 22].

**1.2. Diffusive approximation.** To study diffusive limits of (1.2), one rescales $(t, \mathbf{x}) \to (\varepsilon^2 t, \varepsilon \mathbf{x})$ in order to derive

$$(1.3) \qquad \varepsilon \partial_t f^\pm \pm \partial_x f^\pm = \frac{\sigma(\mathbf{x})}{\varepsilon}\left(\frac{\rho}{4} - f^\pm\right), \qquad \varepsilon \partial_t g^\pm \pm \partial_y g^\pm = \frac{\sigma(\mathbf{x})}{\varepsilon}\left(\frac{\rho}{4} - g^\pm\right)$$

and introduces macroscopic quantities, mass and flux,

$$\rho = f^+ + f^- + g^+ + g^-, \qquad \mathbf{J} = \frac{1}{\varepsilon}\begin{pmatrix} f^+ - f^- \\ g^+ - g^- \end{pmatrix} \in \mathbb{R}^2.$$

By summing the four balance laws, the continuity equation emerges:

$$\partial_t \rho + \operatorname{div} \mathbf{J} = 0.$$

However, as noted in [19, page 504], the equation on $\mathbf{J}$ is not closed,

$$(1.4) \qquad \varepsilon^2 \partial_t \mathbf{J} + \nabla \begin{pmatrix} f^+ + f^- \\ g^+ + g^- \end{pmatrix} = -\sigma(\mathbf{x})\, \mathbf{J},$$

so that, formally, the asymptotic behavior appears to be given by

$$\partial_t \rho = \partial_x \left( \frac{\partial_x (f^+ + f^-)}{\sigma(\mathbf{x})} \right) + \partial_y \left( \frac{\partial_y (g^+ + g^-)}{\sigma(\mathbf{x})} \right).$$

However, by subtracting the first (second) and the third (fourth) balance laws,

$$\varepsilon \partial_t (f^\pm - g^\pm) \pm (\partial_x f^\pm - \partial_y g^\pm) = -\frac{\sigma}{\varepsilon}(f^\pm - g^\pm),$$

we get that $|f^\pm - g^\pm| = O(\varepsilon)$, and so former calculations can be improved into

$$\varepsilon^2 \partial_t \mathbf{J} + \nabla\left(\frac{\rho}{2}\right) = -\sigma \mathbf{J} - \frac{1}{2}\nabla \begin{pmatrix} (f^+ - g^+) + (f^- - g^-) \\ (g^+ - f^+) + (g^- - f^-) \end{pmatrix} = -\sigma \mathbf{J} - O(\varepsilon),$$

which leads to the expected diffusion equation (see also (4.10))

$$(1.5) \qquad \partial_t \rho(t, \mathbf{x}) = \operatorname{div} \left( \frac{\nabla \rho}{2\sigma(\mathbf{x})} \right), \qquad \text{or } \partial_t \rho = \frac{\Delta \rho}{2\sigma} \text{ if } \sigma \text{ is a constant.}$$

These formal arguments were made fully rigorous in [19] when $\sigma$ is a constant.

**1.3. Plan of the paper.** This text follows a similar roadmap as the original article [16], with the supplementary difficulty that every derivation must now be made on two-dimensional kinetic models. To proceed, we recall in section 2 the pioneering results of [4], thanks to which one can deduce, by means of Laplace transforms, kinetic steady-states from harmonic functions. Following ideas of [14, 15], an $S$-matrix is derived, in section 3, from the data of such polynomial kinetic steady-states, yielding a time-marching scheme (3.5), which is able to preserve nontrivial 2D equilibria (see Theorem 3.2). Moreover, the $S$-matrix being doubly stochastic, it is straightforward to show that (3.5) preserves positivity as well as $L^1/L^\infty$ bounds, like its continuous counterpart. Drawing on our paper [17], after a parabolic rescaling of variables, the

$S$-matrix decomposes nicely so as to yield an implicit-explicit (IMEX) scheme (4.2) which relaxes, as $\varepsilon \to 0$, towards (4.10), which is a consistent discretization of (1.5). Most surprisingly, this IMEX time-marching scheme is still endowed with the well-balanced property; see Theorem 4.1. Rigorous proofs are produced in section 5, in particular in Theorem 5.6, where we can see that the multidimensional feature (1.4), raised in [19], has consequences at the numerical level. These bounds are visualized in section 6, where several challenging benchmarks for both (3.5) and (4.4) are tested on a coarse $32 \times 32$ Cartesian grid. Finally, section 7 paves the way for tackling more complex kinetic models, like (7.1), and some early results of [16] are rephrased in the context of $S$-matrices in Appendix A.

## 2. Harmonic stationary distributions.

### 2.1. Harmonic functions and isotropic scattering.

In [4] (see also [10]), the authors explain how to derive an infinity of (explicit) exact steady-states of the following multidimensional kinetic model:

$$\partial_t f(t,\mathbf{x},\mathbf{v}) + \mathbf{v}\cdot\nabla_{\mathbf{x}} f = \int_{\mathbb{S}^1} f(t,\mathbf{x},\mathbf{v}')\frac{\mathrm{d}\mathbf{v}'}{2\pi} - f, \quad \mathbf{x}=(x,y), \quad \mathbf{v}=(\xi,\eta).$$

By virtue of the method of characteristics, long-time asymptotics $t \to +\infty$ satisfy

$$f(\mathbf{x},\mathbf{v}) = \int_0^\infty \exp(-r)\,\rho(\mathbf{x}-r\mathbf{v})\,\mathrm{d}r, \qquad \rho(\mathbf{x}) = \int_{\mathbb{S}^1} f(\mathbf{x},\mathbf{v})\frac{\mathrm{d}\mathbf{v}}{2\pi}, \tag{2.1}$$

which is the Laplace transform of the (oriented) one-dimensional trace of $\rho$ [21],

$$\tilde{\rho}_{\mathbf{x},\mathbf{v}} : \mathbb{R}^+ \ni r \mapsto \rho(\mathbf{x}-r\mathbf{v}), \qquad f(\mathbf{x},\mathbf{v}) = \mathcal{L}_r(\tilde{\rho}_{\mathbf{x},\mathbf{v}})[p=1]. \tag{2.2}$$

Indeed, the method of characteristics yields, for any $\mathbf{v}\in\mathbb{S}^1$,

$$\frac{d}{ds}\Big(f(s,\mathbf{x}+s\mathbf{v},\mathbf{v})\exp(s)\Big) = \rho(s,\mathbf{x}+s\mathbf{v})\exp(s), \quad \rho(s,\mathbf{x}) = \int_{\mathbb{S}^1} f(s,\mathbf{x},\mathbf{v}')\frac{d\mathbf{v}'}{2\pi},$$

so that, by integrating on $s\in(0,t)$,

$$f(t,\mathbf{x},\mathbf{v}) = f(0,\mathbf{x}-t\mathbf{v},\mathbf{v})\exp(-t) + \int_0^t \rho(\mathbf{x}-r\mathbf{v})\exp(-r)\,\mathrm{d}r.$$

Letting $t\to+\infty$, we (formally) deduce the long-time asymptotics (2.1):

$$f(\mathbf{x},\mathbf{v}) = \int_0^\infty \rho(\mathbf{x}-r\mathbf{v})\exp(-r)\,\mathrm{d}r.$$

A Fredholm equation (of the second kind) follows by integrating again in $\mathbf{v}\in\mathbb{S}^1$:

$$\forall \mathbf{x}\in\mathbb{R}^2, \qquad \rho(\mathbf{x}) = \int_0^\infty \exp(-r)\left(\int_{\mathbb{S}^1}\rho(\mathbf{x}-r\mathbf{v})\,\frac{\mathrm{d}\mathbf{v}}{2\pi}\right)\mathrm{d}r. \tag{2.3}$$

At this point, the authors of [4] claim that, as the long-time behavior of the kinetic model is pure diffusion and $\rho$ is a macroscopic quantity, *harmonic functions may induce mesoscopic steady-states by means of* (2.1). Hence, if $\rho$ is a steady-state of diffusion, $\Delta\rho = 0$, and its mean-value property [6, 12, 23] yields

$$\forall r\in\mathbb{R}^+, \qquad \rho(\mathbf{x}) = \int_{\mathbb{S}^1}\rho(\mathbf{x}-r\mathbf{v})\frac{\mathrm{d}\mathbf{v}}{2\pi},$$

so that, by multiplying by $\exp(-r)$ and integrating in $r \in \mathbb{R}^+$,

$$\int_0^\infty \rho(\mathbf{x}) \exp(-r)\,\mathrm{d}r = \rho(\mathbf{x}) = \int_0^\infty \exp(-r) \left( \int_{\mathbb{S}^1} \rho(\mathbf{x} - r\mathbf{v}) \frac{\mathrm{d}\mathbf{v}}{2\pi} \right) \mathrm{d}r$$

holds for any $\mathbf{x} \in \mathbb{R}^2$, so that (2.3) is satisfied, and a class of stationary kinetic densities $f(\mathbf{x}, \mathbf{v})$ can be deduced from (2.1). For instance, harmonic polynomials furnish an infinity of 2D mesoscopic steady-states, which generalize the only two $1, x - v$ (see, e.g., [13, Chap. 9]), which follow from $\rho''(x) = 0$ in one dimension.

**2.2. Kinetic steady-states and harmonic polynomials.** A major result in [4] is that *kinetic stationary solutions* $f(\mathbf{x}, \mathbf{v})$ *can be deduced from macroscopic (i.e., diffusive or harmonic) ones* $\rho(\mathbf{x})$ by means of a Laplace transform of $r \mapsto \rho(\mathbf{x} - r\,\mathbf{v})$,

$$f(\mathbf{x}, \mathbf{v}) = \int_0^\infty \rho(\mathbf{x} - r\,\mathbf{v}) \exp(-r)\,\mathrm{d}r, \qquad \Delta\rho = 0, \tag{2.4}$$

as soon as certain integrability conditions are met (see [4, Theorem A]). Accordingly, in the special case where $\mathbf{x} = x \in \mathbb{R}$ (one space dimension), harmonic solutions of $\mathrm{d}^2\rho/\mathrm{d}x^2 = 0$ reduce to $\{1, x\}$, and it results that, for $v \in \mathbb{R}$,

$$f(x, v) = \int_0^\infty \exp(-r)\,\mathrm{d}r = 1, \qquad f(x, v) = \int_0^\infty (x - rv) \exp(-r)\,\mathrm{d}r = x - v,$$

which are well-known "separated variables Case's solutions" (see [13, eqn. (9.8)]). In more space dimensions, harmonic functions are abundant (any holomorphic function of $z = x + iy \in \mathbb{C}$ furnishes two harmonic ones: its real and imaginary parts), so that (2.4) yields an infinite set of polynomial solutions, being

$$f(\mathbf{x}, \mathbf{v}) = \Big\{ 1,\ \mathbf{x} - \mathbf{v} \in \mathbb{R}^2, \\ xy - (x\eta + y\xi) + 2\xi\eta,\ \frac{x^2 - y^2}{2} - (x\xi - y\eta) + (\xi^2 - \eta^2), \dots \text{etc.} \Big\}; \tag{2.5}$$

see [4, eqn. (2.6)]. The first ones correspond to "dimensional splitting," whereas the last two are truly 2D and "conjugate" in a certain sense (as seen below). These stationary distributions $f(\mathbf{x}, \mathbf{v})$ can be easily retrieved from (2.4) by taking advantage of the expression of harmonic functions in polar coordinates,

$$\rho(x = r\cos\theta, y = r\sin\theta) = a_0 + \sum_{n \in \mathbb{N}_*} (a_n \cos n\theta + b_n \sin n\theta) r^n, \tag{2.6}$$

in which the first basis components are

$$\Big\{ 1,\ x = r\cos\theta,\ y = r\sin\theta,\ x^2 - y^2 = r^2\cos 2\theta,\ 2xy = r^2 \sin 2\theta, \dots \Big\}.$$

These "harmonic steady-states" $f(\mathbf{x}, \mathbf{v})$ follow from Euler's Gamma function,

$$\Gamma(x) = \int_0^\infty \exp(-t)\, t^{x-1} \mathrm{d}t, \qquad \Gamma(n) = (n-1)! \text{ if } n \in \mathbb{N},$$

because, according to (2.1), the polynomial solutions given in (2.6) rewrite as

$$f(\mathbf{x}, \mathbf{v}) = \Big\{ \Gamma(1),\ \Gamma(1)\mathbf{x} - \Gamma(2)\mathbf{v},\ \Gamma(1)xy - \Gamma(2)(x\eta + y\xi) + \Gamma(3)\xi\eta, \dots \Big\}.$$

**2.3. Isotropic scattering and $S$-matrix.** Yet, in order to follow the ideas of [13, Part II], a 4 × 4 $S$-matrix is derived which relates four "outgoing states" to four "incoming states" (which are available data); see Figure 2.1. Such an $S$-matrix can be seen as the restriction to a finite set of velocities of a "continuous stationary scattering operator" $\mathcal{S}$, defined on any circle of radius $R > 0$ as

$$\boxed{\mathcal{S} : f\Big(R(\cos\theta, \sin\theta); -(\cos\theta, \sin\theta)\Big) \mapsto f\Big(R(\cos\theta, \sin\theta); (\cos\theta, \sin\theta)\Big).} \tag{2.7}$$

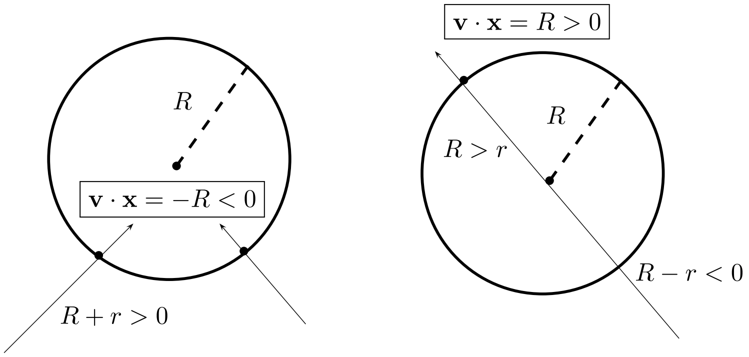
$\mathbf{v}\cdot\mathbf{x} = R > 0$
$R$
$\mathbf{v}\cdot\mathbf{x} = -R < 0$
$R + r > 0$
$R$
$R > r$
$R - r < 0$

FIG. 2.1. *Incoming (left) and outgoing (right) states* $f(\mathbf{x}, \mathbf{v})$.

By definition, we call *an incoming (resp., outgoing) state any state* $f(\mathbf{x}, \mathbf{v})$ *such that* $\mathbf{x}\cdot\mathbf{v} = -R < 0$ *(resp.,* $\mathbf{x}\cdot\mathbf{v} = R > 0$*)*; see again Figure 2.1. By (2.1), they are

$$f\Big(R(\cos\theta, \sin\theta); \mp(\cos\theta, \sin\theta)\Big) = \int_0^\infty \exp(-r)\rho\Big((R \pm r)(\cos\theta, \sin\theta)\Big)\,\mathrm{d}r.$$

Since any macroscopic steady-state $\rho(\mathbf{x})$ is harmonic, in polar coordinates,

$$\begin{aligned}\rho(r,\theta) &= a_0 + \sum_{n\in\mathbb{N}_*} r^n\,(a_n \cos n\theta + b_n \sin n\theta)\\ &= a_0 + a_1\,x + b_1\,y + a_2(x^2 - y^2) + 2b_2\,xy + \cdots,\end{aligned}$$

and this determines the (first four) Fourier coefficients of the resulting $f$,

$$\begin{aligned}f\Big(R(\cos\theta, \sin\theta); \mp(\cos\theta, \sin\theta)\Big) &= \int_0^\infty \exp(-r)\rho\Big((R \pm r)(\cos\theta, \sin\theta)\Big)\,\mathrm{d}r\\ &= \int_0^\infty \exp(-r)\Big[a_0 + (R\pm r)(a_1\cos\theta + b_1\sin\theta) + a_2(R\pm r)^2\cos 2\theta + \cdots\Big]\,\mathrm{d}r\\ &= \Gamma(1)\Big[a_0 + R(a_1\cos\theta + b_1\sin\theta) + R^2 a_2\cos 2\theta\Big]\\ &\qquad \pm\Gamma(2)\Big[(a_1\cos\theta + b_1\sin\theta) + 2Ra_2\cos 2\theta\Big] + \Gamma(3)a_2\cos 2\theta + \cdots\\ &= a_0 + (R\pm 1)(a_1\cos\theta + b_1\sin\theta) + (R^2 \pm 2R + 2)a_2\cos 2\theta + \cdots,\end{aligned} \tag{2.8}$$

$\Gamma(n+1) = n!$ being the Gamma function. Thus, the operator $\mathcal{S}$ acts rather simply in the Fourier space of $2\pi$-periodic functions on the circle (of radius $R$):

- Being $2\pi$-periodic, the incoming state $f(R(\cos\theta, \sin\theta), -(\cos\theta, \sin\theta))$ decomposes as a Fourier series with (real) coefficients $(A_n, B_n)_n$, $n \in \mathbb{N}$.

- By identifying successive Fourier indexes in (2.8), coefficients $(a_n, b_n)_n$ of the incoming state follow from inverting a very simple linear system

$$A_0 = a_0,\ A_1 = (R+1)a_1,\ B_1 = (R+1)b_1,\ A_2 = (R^2+2R+2)a_2.$$

- The first Fourier components of the corresponding outgoing state are

$$\begin{aligned} &f\Big(R(\cos\theta, \sin\theta); (\cos\theta, \sin\theta)\Big) \\ &\qquad = A_0 + \frac{R-1}{1+R}(A_1\cos\theta + B_1\sin\theta) + \frac{2-2R+R^2}{2+2R+R^2}A_2\cos 2\theta + \cdots . \end{aligned} \tag{2.9}$$

## 3. A "truly 2D" approximation of $f(t, \mathbf{x}, \mathbf{v})$.

Working on a uniform Cartesian grid for which $\Delta x = \Delta y$, we mimic notation already used in [3]; see Figure 3.1.

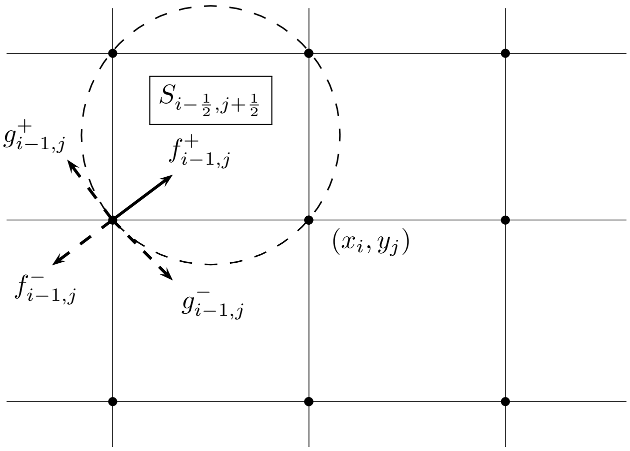

FIG. 3.1. *The S-matrix $S_{i-\frac{1}{2},j+\frac{1}{2}}$ and an incoming state $f^+_{i-1,j}$.*

### 3.1. Derivation of the $S$-matrix.

In order to simulate (1.2) on a 9-point stencil, we only need the first four stationary solutions: the choice between the two "truly 2D" quadratic ones depends on the velocity vectors. A simple case, where one of the conjugate solutions is always null, consists in working in diagonal coordinates,

$$\mathbf{x} = (\mp R, 0) \text{ and } (0, \mp R), \quad \mathbf{v} = (\pm 1, 0) \text{ and } (0, \pm 1), \quad xy - (x\eta + y\xi) + 2\xi\eta \equiv 0,$$

where $R = \Delta x/\sqrt{2}$ is the radius of the disc centered in $x_{i-\frac{1}{2}}, y_{j+\frac{1}{2}}$. As a restriction of (2.7), the $S$-matrix acts on four incoming states and produces four outgoing ones:

$$\begin{pmatrix} f^+_* \\ f^-_* \\ g^+_* \\ g^-_* \end{pmatrix} = S_{i-\frac{1}{2},j+\frac{1}{2}} \begin{pmatrix} f^+_{i-1,j} \\ f^-_{i,j+1} \\ g^+_{i,j} \\ g^-_{i-1,j+1} \end{pmatrix}.$$

By linearity, and following ideas from [13, Chap. 9], a $C^\infty$ stationary solution reads

$$f(\mathbf{x}, \mathbf{v}) = \alpha + \beta(x-\xi) + \gamma(y-\eta) + \nu\left(\frac{x^2-y^2}{2} - (x\xi - y\eta) + (\xi^2 - \eta^2)\right), \tag{3.1}$$

so that the aforementioned incoming and outgoing states are, respectively,

$$\begin{cases} f^+_{i-1,j} = f(\mathbf{x} = (-R, 0), \mathbf{v} = (1, 0)), \quad f^-_{i,j+1} = f(\mathbf{x} = (R, 0), \mathbf{v} = (-1, 0)), \\ g^+_{i,j} = f(\mathbf{x} = (0, -R), \mathbf{v} = (0, 1)), \quad g^-_{i-1,j+1} = f(\mathbf{x} = (0, R), \mathbf{v} = (0, -1)), \end{cases}$$

which, thanks to (3.1), yields a linear system which reads

$$M(\alpha\ \beta\ \gamma\ \nu)^T = (f^+_{i-1,j}\ f^-_{i,j+1}\ g^+_{i,j}\ g^-_{i-1,j+1})^T$$

and

$$\begin{cases} f^+_* = f(\mathbf{x} = (R, 0), \mathbf{v} = (1, 0)), \quad f^-_* = f(\mathbf{x} = (-R, 0), \mathbf{v} = (-1, 0)), \\ g^+_* = f(f\mathbf{x} = (0, R), \mathbf{v} = (0, 1)), \quad g^-_* = f(\mathbf{x} = (0, -R), \mathbf{v} = (0, -1)), \end{cases}$$

involving again the "spectral coefficients" $(\alpha, \beta, \gamma, \nu) \in \mathbb{R}^4$,

$$(f^+_*\ f^-_*\ g^+_*\ g^-_*)^T = \tilde{M}(\alpha\ \beta\ \gamma\ \nu)^T = \tilde{M}\,M^{-1}(f^+_{i-1,j}\ f^-_{i,j+1}\ g^+_{i,j}\ g^-_{i-1,j+1})^T.$$

Accordingly, for an opacity $\sigma \geq 0$, the $S$-matrix decomposes again like

$$\forall (i,j) \in \mathbb{Z}^2, \qquad \boxed{S_{i-\frac{1}{2},j+\frac{1}{2}} = S(\sigma_{i-\frac{1}{2},j+\frac{1}{2}}), \qquad S(\sigma) = \tilde{M}\,M^{-1},} \tag{3.2}$$

where $M$ has mutually orthogonal columns,

$$M = \begin{pmatrix} 1 & -(1+\sigma R) & 0 & 1+(1+\sigma R)^2 \\ 1 & (1+\sigma R) & 0 & 1+(1+\sigma R)^2 \\ 1 & 0 & -(1+\sigma R) & -\left(1+(1+\sigma R)^2\right) \\ 1 & 0 & (1+\sigma R) & -\left(1+(1+\sigma R)^2\right) \end{pmatrix}, \tag{3.3}$$

along with its companion $\tilde{M}$,

$$\tilde{M} = \begin{pmatrix} 1 & -(1-\sigma R) & 0 & 1+(1-\sigma R)^2 \\ 1 & 1-\sigma R & 0 & 1+(1-\sigma R)^2 \\ 1 & 0 & -(1-\sigma R) & -\left(1+(1-\sigma R)^2\right) \\ 1 & 0 & (1-\sigma R) & -\left(1+(1-\sigma R)^2\right) \end{pmatrix},$$

in which a rescaling of $\mathbf{x}$ was made in order to cope with variable opacity $\sigma(\mathbf{x})$. One recognizes the matrices of the 1D Goldstein–Taylor model (see Appendix A and [13, Remark 9.3]),

$$\begin{pmatrix} 1 & -(1+\sigma R) \\ 1 & (1+\sigma R) \end{pmatrix}, \qquad \begin{pmatrix} 1 & -(1-\sigma R) \\ 1 & (1-\sigma R) \end{pmatrix},$$

but now, 1D solutions $\sigma\mathbf{x} - \mathbf{v}$ are coupled by the constant and quadratic ones.

**3.2. Resulting 2D time-marching scheme.** For $\sigma R \geq 0$, the determinant $|M|$ is positive, so $M$ is invertible and its inverse reads

$$|M| = 8(1+\sigma R)^2\left(1+(1+\sigma R)^2\right), \quad M^{-1} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ -A & A & 0 & 0 \\ 0 & 0 & -A & A \\ B & B & -B & -B \end{pmatrix},$$

so that, in (3.1), $\alpha$ is always the average of the four incoming states, and where

$$A = \frac{1}{2(1+\sigma R)}, \qquad B = \frac{1}{4\left(1+(1+\sigma R)^2\right)}.$$

Accordingly, the $S$-matrix is given by the product,

$$\begin{aligned} S(\sigma) &= \tilde{M}\, M^{-1} \\ &= \begin{pmatrix} \frac14 + C + D & \frac14 - C + D & \frac14 - D & \frac14 - D \\ \frac14 - C + D & \frac14 + C + D & \frac14 - D & \frac14 - D \\ \frac14 - D & \frac14 - D & \frac14 + C + D & \frac14 - C + D \\ \frac14 - D & \frac14 - D & \frac14 - C + D & \frac14 + C + D \end{pmatrix}, \end{aligned} \tag{3.4}$$

in which both lines and columns clearly add to unity, because

$$C = \frac{1-\sigma R}{2(1+\sigma R)} = \frac12 - \frac{\sigma R}{1+\sigma R}, \quad D = \frac{(1-\sigma R)^2 + 1}{4((1+\sigma R)^2+1)} = \frac14 - \frac{\sigma R}{1+(1+\sigma R)^2}.$$

The $S$-matrix rewrites as an $O(\sigma R)$-perturbation of the identity of $\mathbb{R}^4$,

$$\begin{aligned} S(\sigma) = \mathrm{Id}_{\mathbb{R}^4} + \sigma R \Bigg\{ & \frac{1}{1+\sigma R} \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \\ & + \frac{1}{1+(1+\sigma R)^2} \begin{pmatrix} -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix} \Bigg\}, \end{aligned}$$

so that, similarly to, e.g., [17, Prop. 3.2],

$$S(\sigma) \to \mathrm{Id}_{\mathbb{R}^4} \text{ if } \sigma \to 0, \qquad S(\sigma) \to S^0 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \text{ if } \sigma \to +\infty.$$

Having at hand the $4 \times 4$ matrix (3.2) allows one to deduce a time-marching scheme for the 2D system (1.2) on a uniform Cartesian grid (see Figure 3.1; $\Delta x = \Delta y$),

$$\begin{pmatrix} f^{+,n+1}_{i,j+1} \\ f^{-,n+1}_{i-1,j} \\ g^{+,n+1}_{i-1,j+1} \\ g^{-,n+1}_{i,j} \end{pmatrix} = \left(1 - \frac{\Delta t}{2R}\right) \begin{pmatrix} f^{+,n}_{i,j+1} \\ f^{-,n}_{i-1,j} \\ g^{+,n}_{i-1,j+1} \\ g^{-,n}_{i,j} \end{pmatrix} + \frac{\Delta t}{2R} S(\sigma_{i-\frac12, j+\frac12}) \begin{pmatrix} f^{+,n}_{i-1,j} \\ f^{-,n}_{i,j+1} \\ g^{+,n}_{i,j} \\ g^{-,n}_{i-1,j+1} \end{pmatrix}. \tag{3.5}$$

High-order time-discretizations (see, e.g., [20]) can easily be applied to (3.5).

LEMMA 3.1. *Under the CFL restriction $\Delta t \leq 2R$, the scheme* (3.5) *is consistent with* (1.2) *and preserves positivity. Moreover, it is conservative and $L^\infty$-bounded.*

*Proof.* Under the aforementioned CFL restriction, (3.5) is a convex combination (as advocated in [14, eqn. (2.2)]); hence it preserves positivity because all the entries of $S(\sigma)$ are nonnegative. Besides, doubly stochastic matrices are such that

$$\forall \vec{v} \in \mathbb{R}^4, \qquad \|S(\sigma)\vec{v}\|_\infty \leq \|\vec{v}\|_\infty, \ \|S(\sigma)\vec{v}\|_1 \leq \|\vec{v}\|_1,$$

which implies that (3.5) is bounded in $L^1$ and $L^\infty$. Consistency is shown for $0 \leq \sigma R \ll 1$ (fine grid); at first order, the expression of the $S$-matrix reduces to

$$\frac{1}{1+(1+\sigma R)^2} \simeq \frac{1}{2(1+\sigma R)}, \quad S(\sigma) = \mathrm{Id}_{\mathbb{R}^4} + \frac{\sigma R}{2(1+\sigma R)} \begin{pmatrix} -3 & 1 & 1 & 1 \\ 1 & -3 & 1 & 1 \\ 1 & 1 & -3 & 1 \\ 1 & 1 & 1 & -3 \end{pmatrix},$$

and inserting this expression into (3.5) yields a consistent approximation of (1.2). □

The scheme (3.5) is able to preserve some nontrivial 2D equilibria; see, e.g., [1].

THEOREM 3.2 (2D well-balanced). *Let $\sigma(\mathbf{x}) \equiv \bar\sigma > 0$ be a constant; then any linear combination* (3.1) *induces a numerical steady-state for the scheme* (3.5),

$$f^{\pm}\left(\frac{x-y}{\sqrt{2}}, \frac{x+y}{\sqrt{2}}\right) = f(\bar\sigma\mathbf{x}; (\pm1, 0)), \qquad g^{\pm}\left(\frac{x-y}{\sqrt{2}}, \frac{x+y}{\sqrt{2}}\right) = f(\bar\sigma\mathbf{x}; (0, \pm1)).$$

*Proof.* Pick $(\alpha, \beta, \gamma, \nu) \in \mathbb{R}^4$ in (3.1), and consider a steady-state $f(\bar\sigma\mathbf{x}, \mathbf{v})$; since $|M| > 0$, its restriction to $\mathbf{v} = \{(\pm1, 0), (0, \pm1)\}$ on a uniform Cartesian grid satisfies

$$\begin{pmatrix}\alpha\\ \beta\\ \gamma\\ \nu\end{pmatrix} = M^{-1}\begin{pmatrix} f^{+,n}_{i-1,j}\\ f^{-,n}_{i,j+1}\\ g^{+,n}_{i,j}\\ g^{-,n}_{i-1,j+1}\end{pmatrix} \quad\Rightarrow\quad S(\bar\sigma)\begin{pmatrix} f^{+,n}_{i-1,j}\\ f^{-,n}_{i,j+1}\\ g^{+,n}_{i,j}\\ g^{-,n}_{i-1,j+1}\end{pmatrix} = \begin{pmatrix} f^{+,n}_{i,j+1}\\ f^{-,n}_{i-1,j}\\ g^{+,n}_{i-1,j+1}\\ g^{-,n}_{i,j}\end{pmatrix},$$

so they are invariant by the time-marching scheme (3.5). By a $-\frac{\pi}{4}$ rotation we pass from diagonal coordinates with $\mathbf{v} = \{(\pm1, 0), (0, \pm1)\}$ to axial ones with (1.1). □

*Remark* 1 (about a full 2D WB property). Given the explicit form of the isotropic scattering operator (2.7), it could be theoretically possible to derive a 2D numerical scheme able to preserve a class of steady-states wider than (3.1), assuming they are band-limited (hence belonging to a Paley–Wiener space). Indeed, under this hypothesis, given a set of pointwise values $(f^{\pm}_{i,j}, g^{\pm}_{i,j})$ of a band-limited steady-state, its corresponding continuous distribution can be recovered thanks to Shannon's sampling theorem. Accordingly, more Fourier coefficients can be identified in (2.8)–(2.9) thanks to the supplementary discrete values available on each circle; see Figure 3.1.

## 4. Diffusive behavior of the $S$-matrix.

In order to study asymptotic limits so as to check consistency with (1.3) and the estimates stated in [19, Theorem 5.1], we rescale $\sigma(\mathbf{x}) \to \sigma(\mathbf{x})/\varepsilon$, $\varepsilon \ll 1$. Accordingly, the $S$-matrix decomposes into

$$S^{\varepsilon} = S^0 + \varepsilon\, S^{1,\varepsilon}, \qquad S^0 = \lim_{\varepsilon\to0} S(\frac{\sigma}{\varepsilon}) = \begin{pmatrix} 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{pmatrix}, \tag{4.1}$$

like in [17, sect. 1.2], so that, following again [17], an IMEX scheme can be deduced,

$$\begin{aligned} &\begin{pmatrix} f^{+,n+1}_{i,j+1}\\ f^{-,n+1}_{i-1,j}\\ g^{+,n+1}_{i-1,j+1}\\ g^{-,n+1}_{i,j}\end{pmatrix} + \frac{\Delta t}{2\varepsilon R}\left\{\begin{pmatrix} f^{+,n+1}_{i,j+1}\\ f^{-,n+1}_{i-1,j}\\ g^{+,n+1}_{i-1,j+1}\\ g^{-,n+1}_{i,j}\end{pmatrix} - S^0\begin{pmatrix} f^{+,n+1}_{i-1,j}\\ f^{-,n+1}_{i,j+1}\\ g^{+,n+1}_{i,j}\\ g^{-,n+1}_{i-1,j+1}\end{pmatrix}\right\}\\ &\qquad = \begin{pmatrix} f^{+,n}_{i,j+1}\\ f^{-,n}_{i-1,j}\\ g^{+,n}_{i-1,j+1}\\ g^{-,n}_{i,j}\end{pmatrix} + \frac{\Delta t}{2R}\, S^{1,\varepsilon}\begin{pmatrix} f^{+,n}_{i-1,j}\\ f^{-,n}_{i,j+1}\\ g^{+,n}_{i,j}\\ g^{-,n}_{i-1,j+1}\end{pmatrix}, \end{aligned} \tag{4.2}$$

and we expect the (implicit, but not costly) left-hand side to yield "Maxwellian estimates" of the type [19, eqn. (5.15)] and the (explicit) right-hand side to produce accurate and consistent diffusive numerical fluxes.

**4.1. Decomposition of the $S$-matrix.** By defining the positive coefficients,

$$\alpha^\varepsilon_{i-\frac12,j+\frac12} = \frac{1}{\varepsilon + \sigma_{i-\frac12,j+\frac12}R}, \qquad \beta^\varepsilon_{i-\frac12,j+\frac12} = \frac{\sigma_{i-\frac12,j+\frac12}R}{\varepsilon^2 + (\varepsilon + \sigma_{i-\frac12,j+\frac12}R)^2}, \tag{4.3}$$

the decomposition (4.1) reads, at each location $i-\frac12, j+\frac12$,

$$S^\varepsilon = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} + \varepsilon \begin{pmatrix} \alpha-\beta & -(\alpha+\beta) & \beta & \beta \\ -(\alpha+\beta) & \alpha-\beta & \beta & \beta \\ \beta & \beta & \alpha-\beta & -(\alpha+\beta) \\ \beta & \beta & -(\alpha+\beta) & \alpha-\beta \end{pmatrix};$$

hence the IMEX scheme (4.2) rewrites as

$$\begin{aligned}
f^{+,n+1}_{i,j+1} &+ \frac{\Delta t}{2\varepsilon R}\Big(f^{+,n+1}_{i,j+1} - f^{-,n+1}_{i,j+1}\Big) = f^{+,n}_{i,j+1} + \\
&\frac{\Delta t}{2R}\Big[\alpha^\varepsilon_{i-\frac12,j+\frac12}\Big(f^{+,n}_{i-1,j} - f^{-,n}_{i,j+1}\Big) + \beta^\varepsilon_{i-\frac12,j+\frac12}\Big(-f^{+,n}_{i-1,j} - f^{-,n}_{i,j+1} + g^{+,n}_{i,j} + g^{-,n}_{i-1,j+1}\Big)\Big] \\
f^{-,n+1}_{i-1,j} &+ \frac{\Delta t}{2\varepsilon R}\Big(f^{-,n+1}_{i-1,j} - f^{+,n+1}_{i-1,j}\Big) = f^{-,n}_{i-1,j} + \\
&\frac{\Delta t}{2R}\Big[\alpha^\varepsilon_{i-\frac12,j+\frac12}\Big(f^{-,n}_{i,j+1} - f^{+,n}_{i-1,j}\Big) + \beta^\varepsilon_{i-\frac12,j+\frac12}\Big(-f^{+,n}_{i-1,j} - f^{-,n}_{i,j+1} + g^{+,n}_{i,j} + g^{-,n}_{i-1,j+1}\Big)\Big] \\
g^{+,n+1}_{i-1,j+1} &+ \frac{\Delta t}{2\varepsilon R}\Big(g^{+,n+1}_{i-1,j+1} - g^{-,n+1}_{i-1,j+1}\Big) = g^{+,n}_{i-1,j+1} + \\
&\frac{\Delta t}{2R}\Big[\alpha^\varepsilon_{i-\frac12,j+\frac12}\Big(g^{+,n}_{i,j} - g^{-,n}_{i-1,j+1}\Big) + \beta^\varepsilon_{i-\frac12,j+\frac12}\Big(f^{+,n}_{i-1,j} + f^{-,n}_{i,j+1} - g^{+,n}_{i,j} - g^{-,n}_{i-1,j+1}\Big)\Big] \\
g^{-,n+1}_{i,j} &+ \frac{\Delta t}{2\varepsilon R}\Big(g^{-,n+1}_{i,j} - g^{+,n+1}_{i,j}\Big) = g^{-,n}_{i,j} + \\
&\frac{\Delta t}{2R}\Big[\alpha^\varepsilon_{i-\frac12,j+\frac12}\Big(g^{-,n}_{i-1,j+1} - g^{+,n}_{i,j}\Big) + \beta^\varepsilon_{i-\frac12,j+\frac12}\Big(f^{+,n}_{i-1,j} + f^{-,n}_{i,j+1} - g^{+,n}_{i,j} - g^{-,n}_{i-1,j+1}\Big)\Big].
\end{aligned}$$

The left-hand side of (4.2) is block-diagonal, so that an index-shift yields

(4.4)

$$\begin{aligned}
&\begin{pmatrix} 1+\frac{\Delta t}{2\varepsilon R} & -\frac{\Delta t}{2\varepsilon R} & 0 & 0 \\ -\frac{\Delta t}{2\varepsilon R} & 1+\frac{\Delta t}{2\varepsilon R} & 0 & 0 \\ 0 & 0 & 1+\frac{\Delta t}{2\varepsilon R} & -\frac{\Delta t}{2\varepsilon R} \\ 0 & 0 & -\frac{\Delta t}{2\varepsilon R} & 1+\frac{\Delta t}{2\varepsilon R} \end{pmatrix} \begin{pmatrix} f^{+,n+1}_{i,j} \\ f^{-,n+1}_{i,j} \\ g^{+,n+1}_{i,j} \\ g^{-,n+1}_{i,j} \end{pmatrix} = \begin{pmatrix} f^{+,n}_{i,j} \\ f^{-,n}_{i,j} \\ g^{+,n}_{i,j} \\ g^{-,n}_{i,j} \end{pmatrix} \\
&+ \frac{\Delta t}{2R} \begin{pmatrix} \alpha^\varepsilon_{i-\frac12,j-\frac12}(f^{+,n}_{i-1,j-1} - f^{-,n}_{i,j}) - \beta^\varepsilon_{i-\frac12,j-\frac12}(f^{+,n}_{i-1,j-1} + f^{-,n}_{i,j} - g^{+,n}_{i,j-1} - g^{-,n}_{i-1,j}) \\ \alpha^\varepsilon_{i+\frac12,j+\frac12}(f^{-,n}_{i+1,j+1} - f^{+,n}_{i,j}) - \beta^\varepsilon_{i+\frac12,j+\frac12}(f^{+,n}_{i,j} + f^{-,n}_{i+1,j+1} - g^{+,n}_{i+1,j} - g^{-,n}_{i,j+1}) \\ \alpha^\varepsilon_{i+\frac12,j-\frac12}(g^{+,n}_{i+1,j-1} - g^{-,n}_{i,j}) + \beta^\varepsilon_{i+\frac12,j-\frac12}(f^{+,n}_{i,j-1} + f^{-,n}_{i+1,j} - g^{+,n}_{i+1,j-1} - g^{-,n}_{i,j}) \\ \alpha^\varepsilon_{i-\frac12,j+\frac12}(g^{-,n}_{i-1,j+1} - g^{+,n}_{i,j}) + \beta^\varepsilon_{i-\frac12,j+\frac12}(f^{+,n}_{i-1,j} + f^{-,n}_{i,j+1} - g^{+,n}_{i,j} - g^{-,n}_{i-1,j+1}) \end{pmatrix}.
\end{aligned}$$

The implicit part relies on a block-diagonal matrix, for which

$$\begin{pmatrix} 1+b & -b \\ -b & 1+b \end{pmatrix}^{-1} = \frac{1}{a+b}\begin{pmatrix} a & b \\ b & a \end{pmatrix}, \quad b = \frac{\Delta t}{2\varepsilon R}, \quad a = 1 + \frac{\Delta t}{2\varepsilon R},$$

so that (4.4) rewrites as an explicit time-marching scheme. The matrix on the left-hand side of (4.4) may be written as

$$\mathrm{Id}_{\mathbb{R}^4} + \frac{\Delta t}{2\varepsilon R}H_0, \qquad \text{with} \qquad H_0 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}. \tag{4.5}$$

Denoting $\mathfrak{f}^n_{i,j} = f^{+,n}_{i,j} + f^{-,n}_{i,j}$ and $\mathfrak{g}^n_{i,j} = g^{+,n}_{i,j} + g^{-,n}_{i,j}$, their time evolution follows from adding the first two and the last two equations in (4.4):

$$\begin{aligned}\mathfrak{f}^{n+1}_{i,j} = {} & \mathfrak{f}^n_{i,j} + \frac{\Delta t}{2R}\Big(\alpha^\varepsilon_{i-\frac12,j-\frac12}(f^{+,n}_{i-1,j-1} - f^{-,n}_{i,j}) + \alpha^\varepsilon_{i+\frac12,j+\frac12}(f^{-,n}_{i+1,j+1} - f^{+,n}_{i,j})\Big) \\ & - \frac{\Delta t}{2R}\Big(\beta^\varepsilon_{i-\frac12,j-\frac12}(f^{+,n}_{i-1,j-1} + f^{-,n}_{i,j} - g^{+,n}_{i,j-1} - g^{-,n}_{i-1,j}) \\ & + \beta^\varepsilon_{i+\frac12,j+\frac12}(f^{+,n}_{i,j} + f^{-,n}_{i+1,j+1} - g^{+,n}_{i+1,j} - g^{-,n}_{i,j+1})\Big),\end{aligned} \tag{4.6}$$

$$\begin{aligned}\mathfrak{g}^{n+1}_{i,j} = {} & \mathfrak{g}^n_{i,j} + \frac{\Delta t}{2R}\Big(\alpha^\varepsilon_{i+\frac12,j-\frac12}(g^{+,n}_{i+1,j-1} - g^{-,n}_{i,j}) + \alpha^\varepsilon_{i-\frac12,j+\frac12}(g^{-,n}_{i-1,j+1} - g^{+,n}_{i,j})\Big) \\ & + \frac{\Delta t}{2R}\Big(\beta^\varepsilon_{i+\frac12,j-\frac12}(f^{+,n}_{i,j-1} + f^{-,n}_{i+1,j} - g^{+,n}_{i+1,j-1} - g^{-,n}_{i,j}) \\ & + \beta^\varepsilon_{i-\frac12,j+\frac12}(f^{+,n}_{i-1,j} + f^{-,n}_{i,j+1} - g^{+,n}_{i,j} - g^{-,n}_{i-1,j+1})\Big).\end{aligned} \tag{4.7}$$

THEOREM 4.1. *The IMEX scheme* (4.2) *is "2D well balanced" in the sense that, if the data are at steady-state for all* $i, j \in \mathbb{Z}^2$, $A^n_{i-\frac12,j+\frac12} = S^\varepsilon_{i-\frac12,j+\frac12}\, B^n_{i-\frac12,j+\frac12}$, *where*

$$A^n_{i-\frac12,j+\frac12} := \begin{pmatrix} f^{+,n}_{i,j+1} \\ f^{-,n}_{i-1,j} \\ g^{+,n}_{i-1,j+1} \\ g^{-,n}_{i,j} \end{pmatrix}, \quad B^n_{i-\frac12,j+\frac12} := \begin{pmatrix} f^{+,n}_{i-1,j} \\ f^{-,n}_{i,j+1} \\ g^{+,n}_{i,j} \\ g^{-,n}_{i-1,j+1} \end{pmatrix},$$

*then, uniformly in* $\varepsilon > 0$, *they are kept invariant:*

$$\boxed{\forall (i,j) \in \mathbb{Z}^2, \qquad f^{\pm,n+1}_{i,j} = f^{\pm,n}_{i,j}, \qquad g^{\pm,n+1}_{i,j} = g^{\pm,n}_{i,j}.} \tag{4.8}$$

*Proof.* Denote $S := S^\varepsilon_{i-\frac12,j+\frac12}$ as $\varepsilon$ is fixed; since $S = S^0 + \varepsilon S^{1,\varepsilon}$, (4.2) rewrites as

$$A^{n+1}_{i-\frac12,j+\frac12} + \frac{\Delta t}{2\varepsilon R}\left(A^{n+1}_{i-\frac12,j+\frac12} - S^0 B^{n+1}_{i-\frac12,j+\frac12}\right) = A^n_{i-\frac12,j+\frac12} + \frac{\Delta t}{2\varepsilon R} \times \varepsilon S^{1,\varepsilon}\, B^n_{i-\frac12,j+\frac12}.$$

Being at steady-state, the data satisfy

$$A^n_{i-\frac12,j+\frac12} = S\, B^n_{i-\frac12,j+\frac12} = (S^0 + \varepsilon S^{1,\varepsilon}) B^n_{i-\frac12,j+\frac12},$$

so the IMEX scheme rewrites, for this particular type of data,

$$\begin{aligned}(A^{n+1}_{i-\frac12,j+\frac12} - A^n_{i-\frac12,j+\frac12}) + \frac{\Delta t}{2\varepsilon R}\left(A^{n+1}_{i-\frac12,j+\frac12} - S^0 B^{n+1}_{i-\frac12,j+\frac12}\right) \\ = \frac{\Delta t}{2\varepsilon R}\left(A^n_{i-\frac12,j+\frac12} - S^0 B^n_{i-\frac12,j+\frac12}\right).\end{aligned}$$

Yet, (4.8) is reached by shifting indexes like in (4.4), along with the following:

- adding both the first and last two equations yields preservation of averages,

$$\mathfrak{f}^{n+1}_{i,j} = \mathfrak{f}^n_{i,j}, \qquad \mathfrak{g}^{n+1}_{i,j} = \mathfrak{g}^n_{i,j}, \qquad \text{and} \qquad \rho^{n+1}_{i,j} = \rho^n_{i,j};$$

- subtracting them, while defining macroscopic fluxes, yields

$$\mathbf{J}^n_{i,j} = \begin{pmatrix} f^+ - f^- \\ g^+ - g^- \end{pmatrix}^n_{i,j} \in \mathbb{R}^2, \qquad \left(1 + \frac{\Delta t}{\varepsilon R}\right)(\mathbf{J}^{n+1}_{i,j} - \mathbf{J}^n_{i,j}) = 0;$$

- and finally taking advantage of

$$\begin{pmatrix} f^\pm \\ g^\pm \end{pmatrix} = \frac12 \left( \begin{pmatrix} \mathfrak{f} \\ \mathfrak{g} \end{pmatrix} \pm \mathbf{J} \right). \qquad \square$$

**4.2. Formal diffusive limit.** When $\varepsilon \to 0$, we deduce from (4.4) that

$$\begin{pmatrix} f_{i,j}^{+,n+1} \\ f_{i,j}^{-,n+1} \\ g_{i,j}^{+,n+1} \\ g_{i,j}^{-,n+1} \end{pmatrix} \in \operatorname{Ker}(H_0) = \operatorname{Span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right\}. \tag{4.9}$$

Then, in the limit $\varepsilon \to 0$, we expect, at least formally, that

$$f_{i,j}^{+,n+1} = f_{i,j}^{-,n+1} = \frac{1}{2}\mathfrak{f}_{i,j}^{n+1}, \qquad g_{i,j}^{+,n+1} = g_{i,j}^{-,n+1} = \frac{1}{2}\mathfrak{g}_{i,j}^{n+1},$$

along with, from (4.3),

$$\alpha^{\varepsilon}_{i-\frac{1}{2},j+\frac{1}{2}}, \beta^{\varepsilon}_{i-\frac{1}{2},j+\frac{1}{2}} \to \frac{1}{\sigma_{i-\frac{1}{2},j+\frac{1}{2}} R}, \qquad \varepsilon \to 0,$$

so that the former equations (4.6) and (4.7) become

$$\begin{aligned}
\mathfrak{f}_{i,j}^{n+1} = \mathfrak{f}_{i,j}^{n} + \frac{\Delta t}{4R^2} \Bigg( & \frac{1}{\sigma_{i-\frac{1}{2},j-\frac{1}{2}}} \Big( (\mathfrak{g}_{i-1,j}^{n} - \mathfrak{f}_{i,j}^{n}) + (\mathfrak{g}_{i,j-1}^{n} - \mathfrak{f}_{i,j}^{n}) \Big) \\
& + \frac{1}{\sigma_{i+\frac{1}{2},j+\frac{1}{2}}} \Big( (\mathfrak{g}_{i+1,j}^{n} - \mathfrak{f}_{i,j}^{n}) + (\mathfrak{g}_{i,j+1}^{n} - \mathfrak{f}_{i,j}^{n}) \Big) \Bigg), \\
\mathfrak{g}_{i,j}^{n+1} = \mathfrak{g}_{i,j}^{n} + \frac{\Delta t}{4R^2} \Bigg( & \frac{1}{\sigma_{i-\frac{1}{2},j+\frac{1}{2}}} \Big( (\mathfrak{f}_{i,j+1}^{n} - \mathfrak{g}_{i,j}^{n}) + (\mathfrak{f}_{i-1,j}^{n} - \mathfrak{g}_{i,j}^{n}) \Big) \\
& + \frac{1}{\sigma_{i+\frac{1}{2},j-\frac{1}{2}}} \Big( (\mathfrak{f}_{i+1,j}^{n} - \mathfrak{g}_{i,j}^{n}) + (\mathfrak{f}_{i,j-1}^{n} - \mathfrak{g}_{i,j}^{n}) \Big) \Bigg).
\end{aligned}$$

Accordingly, $\mathfrak{f}$ and $\mathfrak{g}$ satisfy similar diffusion equations if the opacity $\sigma$ is smooth. Consequently, if initially they are close enough (so-called *well-prepared initial data*), they can be expected to stay so because their difference $\mathfrak{f}_{i,j}^{n} - \mathfrak{g}_{i,j}^{n}$ satisfies a parabolic equation. The decay of $\mathfrak{f} - \mathfrak{g}$ will be rigorously proved when $\sigma$ is a constant; see Theorem 5.6. Adding, assuming $\mathfrak{f} - \mathfrak{g} \to 0$, and denoting $\rho_{i,j}^{n} = \mathfrak{f}_{i,j}^{n} + \mathfrak{g}_{i,j}^{n}$, yields

$$\begin{aligned}
\rho_{i,j}^{n+1} = \rho_{i,j}^{n} + \frac{\Delta t}{4R^2} \Bigg( & \left( \frac{1}{2\sigma_{i+\frac{1}{2},j+\frac{1}{2}}} + \frac{1}{2\sigma_{i+\frac{1}{2},j-\frac{1}{2}}} \right) (\rho_{i+1,j}^{n} - \rho_{i,j}^{n}) \\
& + \left( \frac{1}{2\sigma_{i+\frac{1}{2},j+\frac{1}{2}}} + \frac{1}{2\sigma_{i-\frac{1}{2},j+\frac{1}{2}}} \right) (\rho_{i,j+1}^{n} - \rho_{i,j}^{n}) \\
& - \left( \frac{1}{2\sigma_{i-\frac{1}{2},j+\frac{1}{2}}} + \frac{1}{2\sigma_{i-\frac{1}{2},j-\frac{1}{2}}} \right) (\rho_{i,j}^{n} - \rho_{i-1,j}^{n}) \\
& - \left( \frac{1}{2\sigma_{i+\frac{1}{2},j-\frac{1}{2}}} + \frac{1}{2\sigma_{i-\frac{1}{2},j-\frac{1}{2}}} \right) (\rho_{i,j}^{n} - \rho_{i,j-1}^{n}) \Bigg),
\end{aligned} \tag{4.10}$$

which is a second-order, finite-difference, monotone (under the CFL restriction (5.6)) discretization of the macroscopic diffusion equation (1.5).

## 5. Rigorous diffusive asymptotic-preserving convergence.

Letting $(u_{i,j})$ stand for any real sequence, we introduce the following notation:

$$\begin{aligned}
&\delta u_{i+\frac{1}{2},j} = u_{i+1,j} - u_{i,j}, \quad \delta u_{i,j+\frac{1}{2}} = u_{i,j+1} - u_{i,j},\\
&\|u\|_1 = \sum_{i,j} \Delta x^2 |u_{i,j}|, \qquad TV(u) = \sum_{i,j} \Delta x \big(|\delta u_{i+\frac{1}{2},j}| + |\delta u_{i,j+\frac{1}{2}}|\big),\\
&\|\Delta u\|_1 = \sum_{i,j} |u_{i+1,j} + u_{i,j+1} + u_{i-1,j} + u_{i,j-1} - 4u_{i,j}|.
\end{aligned} \tag{5.1}$$

### 5.1. General properties of the scheme.

The first stepping stone is the definition of a convenient CFL restriction.

LEMMA 5.1. *Assume that there exists $\sigma_{min} > 0$ such that the opacity is such that $0 < \sigma_{min} \le \sigma_{i-\frac{1}{2},j+\frac{1}{2}}$ for all $i, j$. Then, under the CFL condition*

$$\Delta t \le \min\left\{ \frac{2}{3}\sigma_{min} R^2, \frac{R(\varepsilon + \sigma_{min} R)}{2}\left(1 + \sqrt{1 + \frac{8\varepsilon}{\varepsilon + \sigma_{min} R}}\right)\right\}, \tag{5.2}$$

*the IMEX scheme* (4.4) *preserves positivity.*

*Proof.* Inverting the block-diagonal matrix in (4.4) yields the expressions

$$\begin{aligned}
f_{i,j}^{+,n+1} = \frac{1}{2\varepsilon R + 2\Delta t}\Bigg( &\left(2\varepsilon R + \Delta t - \frac{\Delta t^2}{2R}(\alpha^\varepsilon_{i+\frac{1}{2},j+\frac{1}{2}} + \beta^\varepsilon_{i+\frac{1}{2},j+\frac{1}{2}})\right) f_{i,j}^{+,n}\\
&+ \left(\Delta t - \frac{\Delta t}{2R}(2\varepsilon R + \Delta t)(\alpha^\varepsilon_{i-\frac{1}{2},j-\frac{1}{2}} + \beta^\varepsilon_{i-\frac{1}{2},j-\frac{1}{2}})\right) f_{i,j}^{-,n}\\
&+ (2\varepsilon R + \Delta t)\frac{\Delta t}{2R}(\alpha^\varepsilon_{i-\frac{1}{2},j-\frac{1}{2}} - \beta^\varepsilon_{i-\frac{1}{2},j-\frac{1}{2}}) f_{i-1,j-1}^{+,n}\\
&+ \frac{\Delta t^2}{2R}(\alpha^\varepsilon_{i+\frac{1}{2},j+\frac{1}{2}} - \beta^\varepsilon_{i+\frac{1}{2},j+\frac{1}{2}}) f_{i+1,j+1}^{-,n}\\
&+ (2\varepsilon R + \Delta t)\frac{\Delta t}{2R}\beta^\varepsilon_{i-\frac{1}{2},j-\frac{1}{2}}(g_{i,j-1}^{+,n} + g_{i-1,j}^{-,n})\\
&+ \frac{\Delta t^2}{2R}\beta^\varepsilon_{i+\frac{1}{2},j+\frac{1}{2}}(g_{i+1,j}^{+,n} + g_{i,j+1}^{-,n})\Bigg)
\end{aligned} \tag{5.3}$$

and

$$\begin{aligned}
f_{i,j}^{-,n+1} = \frac{1}{2\varepsilon R + 2\Delta t}\Bigg( &\left(\Delta t - (2\varepsilon R + \Delta t)\frac{\Delta t}{2R}(\alpha^\varepsilon_{i+\frac{1}{2},j+\frac{1}{2}} + \beta^\varepsilon_{i+\frac{1}{2},j+\frac{1}{2}})\right) f_{i,j}^{+,n}\\
&+ \left(2\varepsilon R + \Delta t - \frac{\Delta t^2}{2R}(\alpha^\varepsilon_{i-\frac{1}{2},j-\frac{1}{2}} + \beta^\varepsilon_{i-\frac{1}{2},j-\frac{1}{2}})\right) f_{i,j}^{-,n}\\
&+ \frac{\Delta t^2}{2R}(\alpha^\varepsilon_{i-\frac{1}{2},j-\frac{1}{2}} - \beta^\varepsilon_{i-\frac{1}{2},j-\frac{1}{2}}) f_{i-1,j-1}^{+,n}\\
&+ (2\varepsilon R + \Delta t)\frac{\Delta t}{2R}(\alpha^\varepsilon_{i+\frac{1}{2},j+\frac{1}{2}} - \beta^\varepsilon_{i+\frac{1}{2},j+\frac{1}{2}}) f_{i+1,j+1}^{-,n}\\
&+ \frac{\Delta t^2}{2R}\beta_{i-\frac{1}{2},j-\frac{1}{2}}(g_{i,j-1}^{+,n} + g_{i-1,j}^{-,n})\\
&+ (2\varepsilon R + \Delta t)\frac{\Delta t}{2R}\beta^\varepsilon_{i+\frac{1}{2},j+\frac{1}{2}}(g_{i+1,j}^{+,n} + g_{i,j+1}^{-,n})\Bigg),
\end{aligned}$$

along with similar ones for $g_{i,j}^{\pm,n+1}$, too. From (4.3), we obtain

$$\alpha^{\varepsilon}_{i-\frac{1}{2},j+\frac{1}{2}} - \beta^{\varepsilon}_{i-\frac{1}{2},j+\frac{1}{2}} = \frac{\varepsilon^2 + \varepsilon(\varepsilon + \sigma_{i-\frac{1}{2},j+\frac{1}{2}}R)}{(\varepsilon + \sigma_{i-\frac{1}{2},j+\frac{1}{2}}R)(\varepsilon^2 + (\varepsilon + \sigma_{i-\frac{1}{2},j+\frac{1}{2}}R)^2)} \geq 0.$$

Define a (decreasing) function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$:

$$\psi(x) \overset{def}{=} \frac{1}{\varepsilon + x} + \frac{x}{\varepsilon^2 + (\varepsilon + x)^2}, \qquad \psi'(x) \leq 0 \text{ on } (0, +\infty).$$

Then, since

$$\alpha^{\varepsilon}_{i-\frac{1}{2},j+\frac{1}{2}} + \beta^{\varepsilon}_{i-\frac{1}{2},j+\frac{1}{2}} = \psi(\sigma_{i-\frac{1}{2},j+\frac{1}{2}}R),$$

we get the following bound:

$$\alpha^{\varepsilon}_{i-\frac{1}{2},j+\frac{1}{2}} + \beta^{\varepsilon}_{i-\frac{1}{2},j+\frac{1}{2}} \leq \frac{1}{\varepsilon + \sigma_{min}R} + \frac{\sigma_{min}R}{\varepsilon^2 + (\varepsilon + \sigma_{min}R)^2}.$$

Hence $f_{i,j}^{\pm,n+1}, g_{i,j}^{\pm,n+1}$ are nonnegative combinations of previous iterates if

$$\left(\varepsilon + \frac{\Delta t}{2R}\right)\left(\frac{1}{\varepsilon + \sigma_{min}R} + \frac{\sigma_{min}R}{\varepsilon^2 + (\varepsilon + \sigma_{min}R)^2}\right) \leq 1, \tag{5.4}$$

$$\frac{\Delta t^2}{2R}\left(\frac{1}{\varepsilon + \sigma_{min}R} + \frac{\sigma_{min}R}{\varepsilon^2 + (\varepsilon + \sigma_{min}R)^2}\right) \leq 2\varepsilon R + \Delta t. \tag{5.5}$$

Conditions (5.4) and (5.5) are met if and only if

$$\frac{\Delta t}{2R} \leq \sigma_{min}R \frac{\varepsilon^2 + \varepsilon\sigma_{min}R + (\sigma_{min}R)^2}{2\varepsilon^2 + 3\varepsilon\sigma_{min}R + 2(\sigma_{min}R)^2}, \qquad \frac{\Delta t^2}{R(\varepsilon + \sigma_{min}R)} \leq 2\varepsilon R + \Delta t,$$

and these hold as soon as (5.2) does. ◻

*Remark* 2. A sufficient condition for (5.2) is the heat equation's restriction:

$$2\Delta t \leq \sigma_{min}R^2. \tag{5.6}$$

LEMMA 5.2 (conservation). *Let us assume that the initial data are nonnegative and that* (5.2) *holds. Then, the scheme* (4.4) *is bounded in* $L^1$ *and conservative:*

$$\|f^{+,n}\|_1 + \|f^{-,n}\|_1 + \|g^{+,n}\|_1 + \|g^{-,n}\|_1 = \|f^{+,0}\|_1 + \|f^{-,0}\|_1 + \|g^{+,0}\|_1 + \|g^{-,0}\|_1.$$

*Proof.* It suffices to add the lines of (4.4) and to sum over $i$ and $j$. ◻

LEMMA 5.3 ($L^\infty$ bound). *Let initial data satisfy*

$$0 \leq f_{i,j}^{\pm,0} \leq M, \qquad 0 \leq g_{i,j}^{\pm,0} \leq M.$$

*Then, under the CFL condition* (5.2),

$$\forall\, n \in \mathbb{N}, \qquad 0 \leq f_{i,j}^{\pm,n} \leq M, \qquad 0 \leq g_{i,j}^{\pm,n} \leq M.$$

*Proof.* The proof of Lemma 5.1 yields that, under (5.2), $f_{i,j}^{\pm,n+1}$ and $g_{i,j}^{\pm,n+1}$ are convex combinations of previous iterates, giving the announced $L^\infty$ bound. ◻

*Remark* 3. These bounds hold even if the opacity $\sigma$ is not a (positive) constant.

**5.2. Uniform estimates in the case $\boldsymbol{\sigma}$ is constant.** We study rigorously the diffusive limit of (4.4) in order to prove that it is asymptotic-preserving. Recall from (4.3) the coefficients

$$\alpha^\varepsilon = \frac{1}{\varepsilon + \sigma R}, \qquad \beta^\varepsilon = \frac{\sigma R}{\varepsilon^2 + (\varepsilon + \sigma R)^2} \qquad \text{for } \ \sigma \equiv \bar\sigma \in \mathbb{R}^+. \tag{5.7}$$

As $\sigma$ is constant, (4.4) simplifies into

$$\begin{pmatrix} 1+\frac{\Delta t}{2\varepsilon R} & -\frac{\Delta t}{2\varepsilon R} & 0 & 0 \\ -\frac{\Delta t}{2\varepsilon R} & 1+\frac{\Delta t}{2\varepsilon R} & 0 & 0 \\ 0 & 0 & 1+\frac{\Delta t}{2\varepsilon R} & -\frac{\Delta t}{2\varepsilon R} \\ 0 & 0 & -\frac{\Delta t}{2\varepsilon R} & 1+\frac{\Delta t}{2\varepsilon R} \end{pmatrix} \begin{pmatrix} f_{i,j}^{+,n+1} \\ f_{i,j}^{-,n+1} \\ g_{i,j}^{+,n+1} \\ g_{i,j}^{-,n+1} \end{pmatrix} = \begin{pmatrix} f_{i,j}^{+,n} \\ f_{i,j}^{-,n} \\ g_{i,j}^{+,n} \\ g_{i,j}^{-,n} \end{pmatrix} \tag{5.8}$$
$$+ \frac{\Delta t}{2R} \begin{pmatrix} \alpha^\varepsilon(f_{i-1,j-1}^{+,n} - f_{i,j}^{-,n}) - \beta^\varepsilon(f_{i-1,j-1}^{+,n} + f_{i,j}^{-,n} - g_{i,j-1}^{+,n} - g_{i-1,j}^{-,n}) \\ \alpha^\varepsilon(f_{i+1,j+1}^{-,n} - f_{i,j}^{+,n}) - \beta^\varepsilon(f_{i,j}^{+,n} + f_{i+1,j+1}^{-,n} - g_{i+1,j}^{+,n} - g_{i,j+1}^{-,n}) \\ \alpha^\varepsilon(g_{i+1,j-1}^{+,n} - g_{i,j}^{-,n}) + \beta^\varepsilon(f_{i,j-1}^{+,n} + f_{i+1,j}^{-,n} - g_{i+1,j-1}^{+,n} - g_{i,j}^{-,n}) \\ \alpha^\varepsilon(g_{i-1,j+1}^{-,n} - g_{i,j}^{+,n}) + \beta^\varepsilon(f_{i-1,j}^{+,n} + f_{i,j+1}^{-,n} - g_{i,j}^{+,n} - g_{i-1,j+1}^{-,n}) \end{pmatrix}.$$

LEMMA 5.4. *Let $\sigma$ be a positive constant; under the parabolic CFL restriction,*

$$2\,\Delta t < \sigma_{min} R^2, \tag{5.9}$$

*the scheme* (5.8) *is total variation diminishing (TVD):*

$$\begin{aligned} &TV(f^{+,n+1}) + TV(f^{-,n+1}) + TV(g^{+,n+1}) + TV(g^{-,n+1}) \\ &\qquad\qquad \le TV(f^{+,n}) + TV(f^{-,n}) + TV(g^{+,n}) + TV(g^{-,n}). \end{aligned}$$

*Proof.* By linearity, the expression of $f_{i,j}^{+,n+1}$ in (5.3) in the proof of Lemma 5.1 is similar to the ones of $\delta f_{i+\frac12,j}^{+,n}$. Since (5.9) ensures that coefficients are nonnegative, a triangle inequality yields

$$\begin{aligned} |\delta f_{i+\frac12,j}^{+,n+1}| \ \le\ & \frac{1}{2\varepsilon R + 2\Delta t}\Bigg(\bigg(2\varepsilon R + \Delta t - \frac{\Delta t^2}{2R}(\alpha^\varepsilon + \beta^\varepsilon)\bigg)|\delta f_{i+\frac12,j}^{+,n}| \\ & + \bigg(\Delta t - \frac{\Delta t}{2R}(2\varepsilon R + \Delta t)(\alpha^\varepsilon + \beta^\varepsilon)\bigg)|\delta f_{i+\frac12,j}^{-,n}| \\ & + (2\varepsilon R + \Delta t)\frac{\Delta t}{2R}(\alpha^\varepsilon - \beta^\varepsilon)|\delta f_{i-\frac12,j-1}^{+,n}| \\ & + \frac{\Delta t^2}{2R}(\alpha^\varepsilon - \beta^\varepsilon)|\delta f_{i+\frac32,j+1}^{-,n}| \\ & + (2\varepsilon R + \Delta t)\frac{\Delta t}{2R}\beta^\varepsilon(|\delta g_{i+\frac12,j-1}^{+,n}| + |\delta g_{i-\frac12,j}^{-,n}|) \\ & + \frac{\Delta t^2}{2R}\beta^\varepsilon(|\delta g_{i+\frac12,j}^{+,n}| + |\delta g_{i+\frac12,j+1}^{-,n}|)\Bigg), \end{aligned}$$

with similar expressions for $|\delta f^{-,n}_{i+\frac12,j}|$, $|\delta g^{+,n+1}_{i+\frac12,j}|$, and $|\delta g^{-,n+1}_{i+\frac12,j}|$. Adding

$$\begin{aligned}
&|\delta f^{+,n+1}_{i+\frac12,j}| + |\delta f^{-,n+1}_{i+\frac12,j}| + |\delta g^{+,n+1}_{i+\frac12,j}| + |\delta g^{-,n+1}_{i+\frac12,j}| \\
&\quad \le \left(1 - \frac{\Delta t}{2R}(\alpha^\varepsilon + \beta^\varepsilon)\right)\big(|\delta f^{+,n}_{i+\frac12,j}| + |\delta f^{-,n}_{i+\frac12,j}| + |\delta g^{+,n}_{i+\frac12,j}| + |\delta g^{-,n}_{i+\frac12,j}|\big) \\
&\quad + \frac{\Delta t}{2R}(\alpha^\varepsilon - \beta^\varepsilon)\big(|\delta f^{+,n}_{i-\frac12,j-1}| + |\delta f^{-,n}_{i+\frac32,j+1}| + |\delta g^{+,n}_{i+\frac32,j-1}| + |\delta g^{-,n}_{i-\frac12,j+1}|\big) \\
&\quad + \frac{\Delta t}{2R}\beta^\varepsilon\big(|\delta f^{+,n}_{i+\frac12,j-1}| + |\delta f^{-,n}_{i+\frac32,j}| + |\delta f^{+,n}_{i-\frac12,j}| + |\delta f^{-,n}_{i+\frac12,j+1}|\big) \\
&\quad + \frac{\Delta t}{2R}\beta^\varepsilon\big(|\delta g^{+,n}_{i+\frac12,j-1}| + |\delta g^{-,n}_{i-\frac12,j}| + |\delta g^{+,n}_{i+\frac32,j}| + |\delta g^{-,n}_{i+\frac12,j+1}|\big)
\end{aligned}$$

and summing over $i$ and $j$, we get, after shifting the indexes,

$$\begin{aligned}
&\sum_{i,j}\left(|\delta f^{+,n+1}_{i+\frac12,j}| + |\delta f^{-,n+1}_{i+\frac12,j}| + |\delta g^{+,n+1}_{i+\frac12,j}| + |\delta g^{-,n+1}_{i+\frac12,j}|\right) \\
&\qquad\qquad \le \sum_{i,j}\left(|\delta f^{+,n}_{i+\frac12,j}| + |\delta f^{-,n}_{i+\frac12,j}| + |\delta g^{+,n}_{i+\frac12,j}| + |\delta g^{-,n}_{i+\frac12,j}|\right).
\end{aligned}$$

By the same token, with variations in $j$ instead of $i$, we get the claimed result. ◻

Define $f^\pm_{\Delta x}, g^\pm_{\Delta x}$ as the piecewise constant functions such that

$$f^\pm(t,\mathbf{x}) = f^{\pm,n}_{i,j}, \qquad g^\pm(t,\mathbf{x}) = g^{\pm,n}_{i,j} \tag{5.10}$$

for $t \in [n\Delta t, (n+1)\Delta t)$, $\mathbf{x} \in ((i-\frac12)\Delta x, (i+\frac12)\Delta x) \times ((j-\frac12)\Delta x, (j+\frac12)\Delta x)$.

Corollary 5.5. *Under* (5.9)*, and for bounded integrable nonnegative data, the approximate solutions* (5.10) *are uniformly bounded in* $L^1 \cap L^\infty \cap BV([0,T]\times\mathbb{R}^2)$.

### 5.3. Rigorous diffusive limit with constant opacity.

We are now in position to state the main result of this section.

Theorem 5.6 (asymptotic-preserving property). *Assume* (5.9) *holds and that initial data are independent of* $\varepsilon$ *and smooth enough such that*

$$\exists C \in \mathbb{R}^+, \qquad \|\Delta f^{+,0}\|_1 + \|\Delta f^{-,0}\|_1 + \|\Delta g^{+,0}\|_1 + \|\Delta g^{-,0}\|_1 \le C;$$

*then the sequences* $(f^{\varepsilon}{}^{\pm,n}_{i,j})_\varepsilon$ *and* $(g^{\varepsilon}{}^{\pm,n}_{i,j})_\varepsilon$ *are of uniformly bounded total variation and converge towards limits, denoted, respectively,* $(f^{\pm,n}_{i,j})$ *and* $(g^{\pm,n}_{i,j})$*, which satisfy*

$$f^{+,n}_{i,j} = f^{-,n}_{i,j} = \frac12 \mathfrak{f}^n_{i,j}, \qquad g^{+,n}_{i,j} = g^{-,n}_{i,j} = \frac12 \mathfrak{g}^n_{i,j},$$

*where*

$$\mathfrak{f}^{n+1}_{i,j} = \mathfrak{f}^n_{i,j} + \frac{\Delta t}{4\sigma R^2}(\mathfrak{g}^n_{i,j-1} + \mathfrak{g}^n_{i-1,j} + \mathfrak{g}^n_{i+1,j} + \mathfrak{g}^n_{i,j+1} - 4\mathfrak{f}^n_{i,j}), \tag{5.11}$$

$$\mathfrak{g}^{n+1}_{i,j} = \mathfrak{g}^n_{i,j} + \frac{\Delta t}{4\sigma R^2}(\mathfrak{f}^n_{i,j-1} + \mathfrak{f}^n_{i+1,j} + \mathfrak{f}^n_{i-1,j} + \mathfrak{f}^n_{i,j+1} - 4\mathfrak{g}^n_{i,j}). \tag{5.12}$$

*Moreover, the "Maxwellian gap" decreases in time according to*

$$\forall n \in \mathbb{N}_*, \qquad \|\mathfrak{f}^n - \mathfrak{g}^n\|_1 \le \|\mathfrak{f}^0 - \mathfrak{g}^0\|_1 \exp\left(-\frac{2n\Delta t}{\sigma R^2}\right) + C\,R^2. \tag{5.13}$$

Adding both equations (5.11)–(5.12), we deduce the following result.

COROLLARY 5.7. *Under the same assumptions as Theorem* 5.6, *we have*

$$\rho^{n+1}_{i,j} = \rho^n_{i,j} + \frac{\Delta t}{4\sigma R^2}(\rho^n_{i,j-1} + \rho^n_{i,j+1} + \rho^n_{i-1,j} + \rho^n_{i+1,j} - 4\rho^n_{i,j}), \quad \rho^n = \mathfrak{f}^n + \mathfrak{g}^n,$$

*along with* $f^{\pm,n} = \rho^n/4 + O(R^2)$, $g^{\pm,n} = \rho^n/4 + O(R^2)$.

*Proof.* By the computations in the proof of Lemma 5.4, the sequences $(f^{\varepsilon \pm,n}_{i,j})$ and $(g^{\varepsilon \pm,n}_{i,j})$ are Cauchy sequences with respect to $\varepsilon$ in $\ell^1$. Thus, when $\varepsilon \to 0$, they converge to some limits denoted as, respectively, $(f^{\pm,n}_{i,j})$ and $(g^{\pm,n}_{i,j})$, and we can pass to the limit in (5.8). Hence, as $\varepsilon \to 0$, by (4.5) and (4.9), we get that

$$\forall (i,j,n) \in \mathbb{Z}^2 \times \mathbb{N}, \qquad f^{+,n+1}_{i,j} = f^{-,n+1}_{i,j}, \quad g^{+,n+1}_{i,j} = g^{-,n+1}_{i,j}.$$

Denoting $\mathfrak{f}^{\varepsilon n}_{i,j} = f^{\varepsilon +,n}_{i,j} + f^{\varepsilon -,n}_{i,j}$ and $\mathfrak{g}^{\varepsilon n}_{i,j} = g^{\varepsilon +,n}_{i,j} + g^{\varepsilon -,n}_{i,j}$, we obtain their equations by adding the first two and the last two lines in (5.8):

$$\begin{aligned}\mathfrak{f}^{\varepsilon n+1}_{i,j} = {} & \mathfrak{f}^{\varepsilon n}_{i,j} + \frac{\Delta t}{2R}\Big(\alpha^\varepsilon(f^{\varepsilon +,n}_{i-1,j-1} - f^{\varepsilon -,n}_{i,j}) + \alpha^\varepsilon(f^{\varepsilon -,n}_{i+1,j+1} - f^{\varepsilon +,n}_{i,j})\Big)\\ & - \frac{\Delta t}{2R}\Big(\beta^\varepsilon(f^{\varepsilon +,n}_{i-1,j-1} + \mathfrak{f}^{\varepsilon}_{i,j}{}^{-,n} - g^{\varepsilon +,n}_{i,j-1} - g^{\varepsilon -,n}_{i-1,j})\\ & \quad + \beta^\varepsilon(f^{\varepsilon +,n}_{i,j} + f^{\varepsilon -,n}_{i+1,j+1} - g^{\varepsilon +,n}_{i+1,j} - g^{\varepsilon -,n}_{i,j+1})\Big);\end{aligned} \tag{5.14}$$

$$\begin{aligned}\mathfrak{g}^{\varepsilon n+1}_{i,j} = {} & \mathfrak{g}^{\varepsilon n}_{i,j} + \frac{\Delta t}{2R}\Big(\alpha^\varepsilon(g^{\varepsilon +,n}_{i+1,j-1} - g^{\varepsilon -,n}_{i,j}) + \alpha^\varepsilon(g^{\varepsilon -,n}_{i-1,j+1} - g^{\varepsilon +,n}_{i,j})\Big)\\ & + \frac{\Delta t}{2R}\Big(\beta^\varepsilon(f^{\varepsilon +,n}_{i,j-1} + f^{\varepsilon -,n}_{i+1,j} - g^{\varepsilon +,n}_{i+1,j-1} - g^{\varepsilon -,n}_{i,j})\\ & \quad + \beta^\varepsilon(f^{\varepsilon +,n}_{i-1,j} + f^{\varepsilon -,n}_{i,j+1} - g^{\varepsilon +,n}_{i,j} - g^{\varepsilon -,n}_{i-1,j+1})\Big).\end{aligned} \tag{5.15}$$

From the expressions (5.7),

$$\alpha^\varepsilon, \ \beta^\varepsilon \to \frac{1}{\sigma R} \ \text{ when } \varepsilon \to 0.$$

Yet, passing to the limit, we obtain both (5.11) and (5.12). If initially $\mathfrak{f}$ and $\mathfrak{g}$ are identical, they stay so. More precisely, let $D^n_{i,j} = \mathfrak{f}^n_{i,j} - \mathfrak{g}^n_{i,j}$ be the Maxwellian gap:

$$D^{n+1}_{i,j} = D^n_{i,j}\left(1 - \frac{2\Delta t}{\sigma R^2}\right) + \frac{\Delta t}{4\sigma R^2}(4D^n_{i,j} - D^n_{i,j-1} - D^n_{i-1,j} - D^n_{i+1,j} - D^n_{i,j+1}). \tag{5.16}$$

Hypotheses on initial data in Theorem 5.6 ensure that

$$\|\Delta \mathfrak{f}^0\|_1 + \|\Delta \mathfrak{g}^0\|_1 \le C.$$

Moreover, from (5.11)–(5.12) and (5.9), we have

$$\|\Delta \mathfrak{f}^{n+1}\|_1 \le \|\Delta \mathfrak{f}^n\|_1\left(1 - \frac{\Delta t}{\sigma R^2}\right) + \frac{\Delta t}{\sigma R^2}\|\Delta \mathfrak{g}^n\|_1,$$

$$\|\Delta \mathfrak{g}^{n+1}\|_1 \le \|\Delta \mathfrak{g}^n\|_1\left(1 - \frac{\Delta t}{\sigma R^2}\right) + \frac{\Delta t}{\sigma R^2}\|\Delta \mathfrak{f}^n\|_1.$$

As a consequence, for all $n \in \mathbb{N}$, we have $\|\Delta \mathfrak{f}^n\|_1 + \|\Delta \mathfrak{g}^n\|_1 \le C$, so that

$$\sum_{i,j} |4D^n_{i,j} - D^n_{i,j-1} - D^n_{i-1,j} - D^n_{i+1,j} - D^n_{i,j+1}| \le C.$$

By inserting this latter inequality into (5.16), taking the modulus, and summing,

$$\|D^{n+1}\|_1 = \sum_{i,j} \Delta x^2 |D^{n+1}_{i,j}| \le \|D^n\|_1 \left(1 - \frac{2\Delta t}{\sigma R^2}\right) + C\frac{\Delta t}{\sigma}$$

holds for some constant $C \ge 0$. Applying a discrete Gronwall inequality yields

$$\begin{aligned} \|D^n\|_1 &\le \|D^0\|_1 e^{-2n\Delta t/(\sigma R^2)} + C\frac{\Delta t}{\sigma}\sum_{k=0}^{n-1}\left(1 - \frac{2\Delta t}{\sigma R^2}\right)^k \\ &\le \|D^0\|_1 e^{-2n\Delta t/(\sigma R^2)} + \frac{C}{2}R^2. \end{aligned}$$ ☐

*Remark* 4. The bound (5.13) relates to (1.4) and means that, for constant opacity, $\|\mathfrak{f} - \mathfrak{g}\|_1$ is roughly of order $\Delta x^2$ when nonnegative initial data belong to $W^{2,1}(\mathbb{R}^2)$. Conversely, both $\|f^+ - f^-\|_1$ and $\|g^+ - g^-\|_1$ are of order $\varepsilon$, as in the 1D case; see [13, Lemma 8.4] and [16]. All in all, these will be similar when $\varepsilon \simeq O(R^2)$.

## 6. Numerical assessments.

Hereafter, some benchmarks for both (3.5) and (4.4) are presented on a coarse $32 \times 32$ uniform Cartesian grid. The computational domain is the unit square $\Omega = (0,1)^2$ with various boundary conditions.

### 6.1. Hyperbolic/kinetic scaling: Scheme (3.5).

**6.1.1. Stiff and discontinuous opacity.** Following [9, sect. 5.1], long-time stabilization of (1.2) is considered in the presence of

$$\sigma(\mathbf{x}) = 5 + 995 \cdot \chi\left(\max\left(\left|x - \frac{1}{2}\right|, \left|y - \frac{1}{2}\right|\right) < \frac{1}{4}\right),$$

with $\chi(A)$ the indicator function of a set $A$. Null initial data and an inflow boundary condition are prescribed on the left-hand side by means of

$$f^+(x=0,\cdot) = g^-(x=0,\cdot) = 1,$$

along with specular reflection on horizontal walls $y = 0$, $y = 1$, and outflow at $x = 1$. The macroscopic velocity field $\vec{v}(t,\mathbf{x})$ is defined as the following ratio:

$$\forall \mathbf{x} \in \Omega, \qquad \vec{v}(t,\mathbf{x}) = \begin{pmatrix} \frac{f^+(t,\mathbf{x}) - f^-(t,\mathbf{x})}{\rho(t,\mathbf{x})} \\ \frac{g^+(t,\mathbf{x}) - g^-(t,\mathbf{x})}{\rho(t,\mathbf{x})} \end{pmatrix}, \qquad \text{where } \rho \ne 0.$$

The scheme (3.5) was iterated until $T = 35$ with $\Delta t = 0.975\sqrt{2}\,\Delta x$; see Figure 6.1.

**6.1.2. Smooth, but quickly varying opacity.** Following [7, sect. 4.2], let

$$\begin{aligned} \sigma(\mathbf{x}) = 5 + 195 \exp&\left(-\gamma\left(\left(x - \frac{1}{4}\right)^2 + \left(x - \frac{1}{2}\right)^2 + \left(x - \frac{3}{4}\right)^2\right)\right) \\ &\times \exp\left(-\gamma\left(\left(y - \frac{1}{4}\right)^2 + \left(y - \frac{1}{2}\right)^2 + \left(y - \frac{3}{4}\right)^2\right)\right), \qquad \gamma = 400, \end{aligned}$$

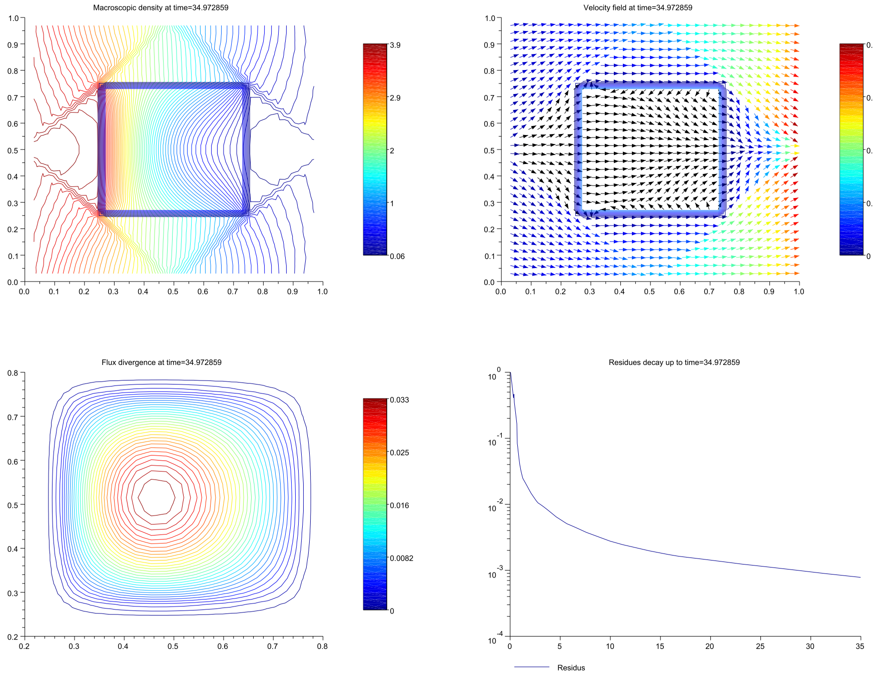

FIG. 6.1. *Steady-state of* (3.5) *in the presence of a square opaque zone.*

with identical initial and boundary conditions. Results are displayed in Figure 6.2. In particular, in both Figures 6.1 and 6.2, a (second-order) centered approximation of the divergence of the macroscopic flux div $\mathbf{J}(t, \mathbf{x})$ is displayed so as to shed light on the ability of (3.5) to stabilize on a correct discretization of stationary regimes. Thanks to the opacity's smoothness, convergence properties for (3.5) can be measured at, say, time $T = 3$ with various grid parameters $\Delta x > 0$; see Table 6.1. More precisely, the output of (3.5) obtained at $T = 3$ with $128 \times 128$ grid points was considered a "reference solution": $\bar{f}^{\pm}, \bar{g}^{\pm}$. Errors $E_{p,N}$ and convergence rates $r_{p,N}$ with respect to $\Delta x > 0$ and $p \geq 1$ can be studied by means of the usual formula,

$$\begin{aligned} E_{p,N} &= \|\bar{f}^{\pm} - f_N^{\pm}\|_{L^p} + \|\bar{g}^{\pm} - g_N^{\pm}\|_{L^p} \\ &= \left(\Delta x^2 \, \|\bar{f}^{\pm} - f_N^{\pm}\|_{\ell^p}^p\right)^{\frac{1}{p}} + \left(\Delta x^2 \, \|\bar{g}^{\pm} - g_N^{\pm}\|_{\ell^p}^p\right)^{\frac{1}{p}}, \\ r_{p,N} &= \frac{\log(E_{p,N}/E_{p,M})}{\log(M/N)}, \end{aligned}$$

where $N, M$ stand for different numbers of grid points.

### 6.2. Diffusive/parabolic scaling: Scheme (4.4).

**6.2.1. Array of opaque Gaussian bumps.** In order to validate the scheme (4.4), the same test was set, along with the parameter $\varepsilon = 10^{-5}$, outflow boundary conditions on each side, and Maxwellian (well-prepared) initial data:

$$\begin{aligned} \rho(t = 0, \mathbf{x}) = \exp&\left(-\nu((x - 0.375)^2 + (x - 0.625)^2)\right) \\ &\times \exp\left(-\nu((y - 0.375)^2 + (y - 0.635)^2)\right), \qquad \nu = 250. \end{aligned}$$

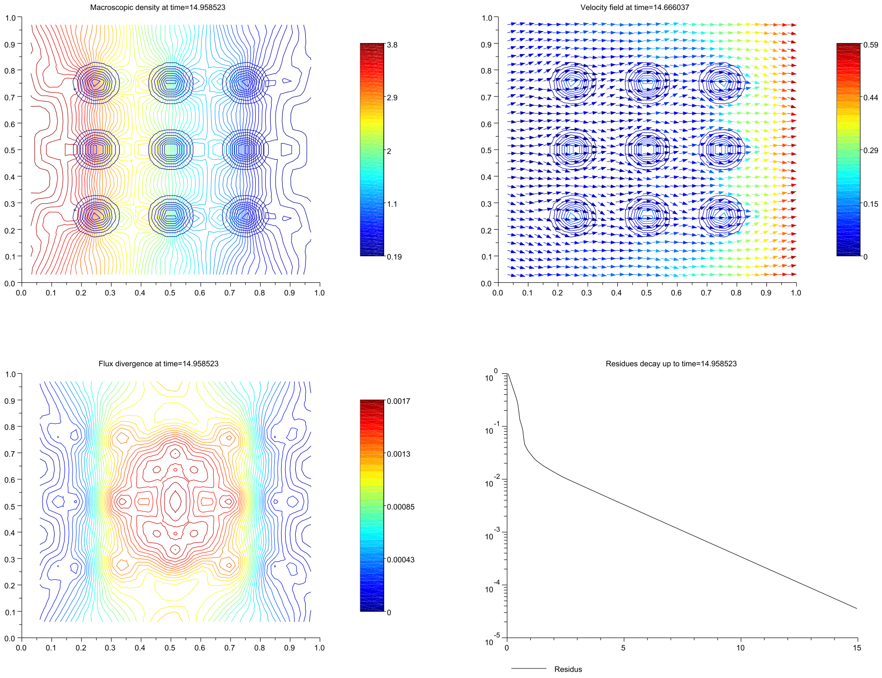

FIG. 6.2. *Steady-state of* (3.5) *in a periodic array of obstacles.*

TABLE 6.1
*Measured convergence rates for* (3.5) *at* $T = 3$.

| Grids | $L^2$ error | $L^\infty$ error | $L^2$ rate | $L^\infty$ rate |
|---|---|---|---|---|
| $N = 26$ | 0.1300071 | 0.6154416 | | |
| $N = 32$ | 0.0860855 | 0.3813428 | 2.0543515 | 2.3852074 |
| $N = 43$ | 0.0427195 | 0.2559275 | 2.4356234 | 1.3862683 |
| $N = 64$ | 0.0175591 | 0.1294132 | 2.2785532 | 1.7475336 |

The scheme was iterated until $T = 0.15$ with the CFL condition (5.6); see Figure 6.3. The macroscopic density is correctly confined inside the array of obstacles, showing how tiny the artificial dissipation of the IMEX scheme really is. The Maxwellian gap $|\mathfrak{f} - \mathfrak{g}|$ is locally of $10^{-3}$, a value compatible with (5.13) because $\Delta x^2 \simeq 10^{-3}$, even in the vicinity of areas of strong variations of $\sigma(\mathbf{x})$; it smoothly decays with time. In addition, $\Delta x^2$ is also the order of accuracy for the centered discretization of the diffusion equation (4.10). The macroscopic velocity field $\vec{v}$ is now rescaled:

$$\forall \mathbf{x} \in \Omega, \qquad \vec{v}(t, \mathbf{x}) = \frac{1}{\varepsilon} \begin{pmatrix} \frac{f^+(t,\mathbf{x}) - f^-(t,\mathbf{x})}{\rho(t,\mathbf{x})} \\ \frac{g^+(t,\mathbf{x}) - g^-(t,\mathbf{x})}{\rho(t,\mathbf{x})} \end{pmatrix}, \qquad \rho \neq 0.$$

### 6.2.2. Gaussian opacity.

A simpler benchmark consists in iterating (4.4) with

$$\forall \mathbf{x} \in \Omega, \qquad \sigma(\mathbf{x}) = 5 + 15 \exp\left( -25 \left( \left| x - \frac{1}{2} \right|^2 + \left| y - \frac{1}{2} \right|^2 \right) \right),$$

with identical initial and outflow boundary conditions, up to $T = 0.1$; see Figure 6.4.

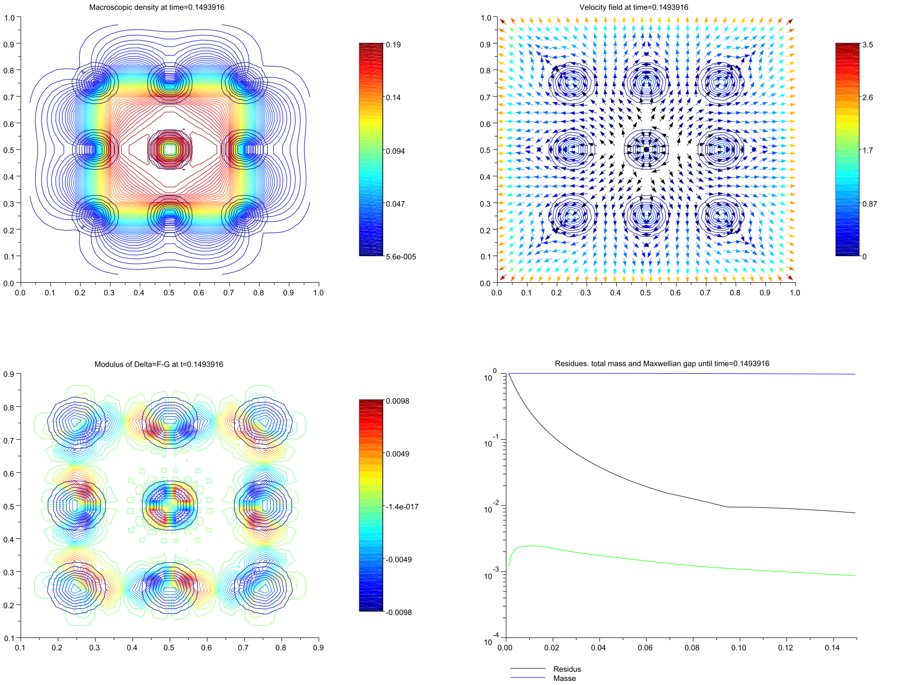

Fig. 6.3. *Diffusive approximation* (4.4) *at* $T = 0.15$, $\varepsilon = 10^{-5}$ *in a periodic array of obstacles.*

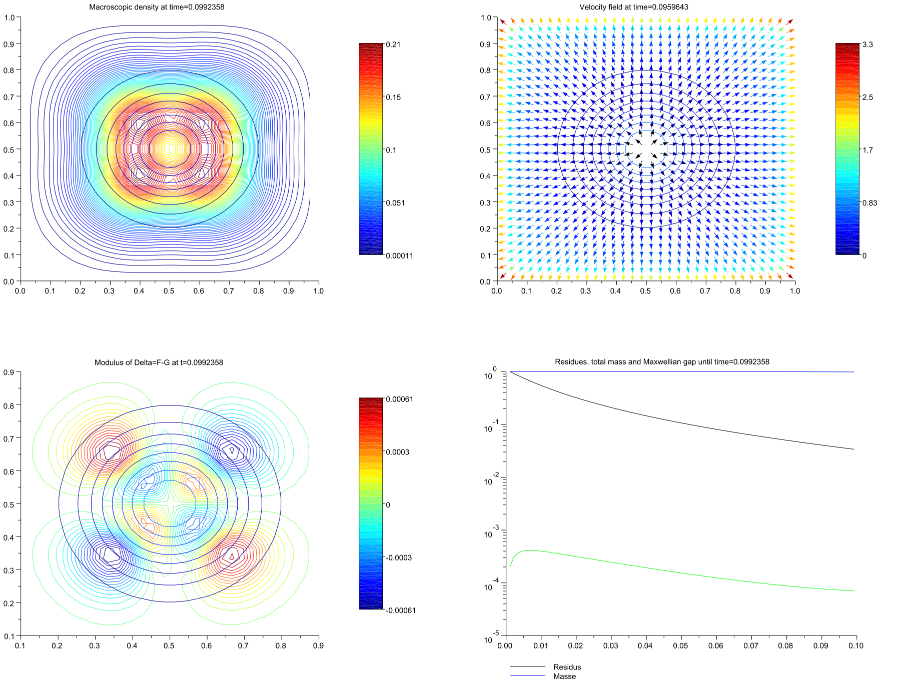

Fig. 6.4. *Diffusive approximation* (4.4) *at* $T = 0.1$, $\varepsilon = 10^{-5}$ *with a Gaussian opacity.*

## 7. Conclusion and outlook.

The present paper showed that a genuinely two-dimensional, numerical scheme (3.5), (4.4) can be deduced from the computations achieved in [4]. Such a strategy is limited neither to isotropic scattering nor to two-dimensional problems.

- Indeed, the idea in [4] works in any space dimension and allows one to derive kinetic steady-states from macroscopic (harmonic) ones. An $8 \times 8$ scattering $S$-matrix follows, and in 3D, it will relate outgoing states to incoming ones on spheres $\mathbb{S}^2$ of radius $R$. The 3D analogue of (3.5) will involve a 27-point stencil and replace, e.g., (2.6) by a spherical harmonics expansion.
- Following [13, sect. 10.3], an elementary model of chemotaxis dynamics is

$$\partial_t f + \mathbf{v} \cdot \nabla f = \chi(\mathbf{v} \cdot \nabla S)\rho(t,\mathbf{x}) - f(\mathbf{x},\mathbf{v}), \qquad \mathbf{a} := \nabla S(\bar{\mathbf{x}}) \in \mathbb{R}^2, \tag{7.1}$$

where $\nabla S$ is "frozen" locally at a point $\bar{\mathbf{x}}$ and the biasing function $\chi \geq 0$ is normalized so as to get the standard 2D continuity equation

$$\int_{\mathbb{S}^1} \chi(\mathbf{v}) \frac{\mathrm{d}\mathbf{v}}{2\pi} = 1, \qquad \partial_t \rho(t,\mathbf{x}) + \operatorname{div}\, \mathbf{J} = 0.$$

The analogue of the Laplace transform in (2.1) for steady-states $f(\mathbf{x},\mathbf{v})$ reads

$$f(\mathbf{x},\mathbf{v}) = \chi(\mathbf{v}) \int_0^\infty \exp(-r)\rho(\mathbf{x} - r\mathbf{v})\,\mathrm{d}r = \chi(\mathbf{v})\mathcal{L}_r(\tilde{\rho}_{\mathbf{x},\mathbf{v}})[p=1], \tag{7.2}$$

which yields a new Fredholm equation, now involving the biasing function $\chi$,

$$\rho(\mathbf{x}) = \int_0^\infty \exp(-r) \left( \int_{\mathbb{S}^1} \chi(\mathbf{v})\, \rho(\mathbf{x} - r\mathbf{v}) \frac{\mathrm{d}\mathbf{v}}{2\pi} \right) \mathrm{d}r. \tag{7.3}$$

To mimic some computations of [4], macroscopic steady-states should verify

$$\forall r \in \mathbb{R}^+, \qquad \rho(\mathbf{x}) = \int_{\mathbb{S}^1} \chi(\mathbf{v})\rho(\mathbf{x} - r\mathbf{v}) \frac{\mathrm{d}\mathbf{v}}{2\pi},$$

which means that *our biasing function $\chi$ should also be the Poisson kernel of a certain elliptic differential operator that $\rho(\mathbf{x})$ solves.* Indeed, $\rho(\mathbf{x} - r\mathbf{v})$ is "boundary data" on $\mathbb{S}^1$, so $\rho(\mathbf{x})$ is the "solution value." Accordingly, from [3, eqn. (2.24)],

$$\rho(\mathbf{x}) = \int_0^{2\pi} \rho\left(\mathbf{x} + r\, e^{i\theta}\right) \frac{\exp(-\omega\, r \cos(\theta - \mu))}{\mathcal{I}_0(\omega\, r)} \frac{\mathrm{d}\theta}{2\pi},$$

so that, by changing $\theta \to \theta - \pi$, one gets (see also [6, 12, 23])

$$\rho(\mathbf{x}) = \int_{-\pi}^{\pi} \rho\left(\mathbf{x} - r\, e^{i\theta}\right) \frac{\exp(\omega\, r \cos(\theta - \mu))}{\mathcal{I}_0(\omega\, r)} \frac{\mathrm{d}\theta}{2\pi}.$$

Pick, as the biasing function ($\mathcal{I}_0$, the modified Bessel function of index zero),

$$\chi(\mathbf{v} = e^{i\theta}) = \frac{\exp(\omega\, r \cos(\theta - \mu))}{\mathcal{I}_0(\omega\, r)} \geq 0, \qquad \int_0^{2\pi} \frac{\exp(\omega\, r \cos\theta)}{\mathcal{I}_0(\omega\, r)} \frac{\mathrm{d}\theta}{2\pi} = 1,$$

the normalization being a consequence of the "integral representation of Bessel functions" (see, e.g., [3, eqn. (3.1)]). Then such a kernel corresponds to drift-diffusion equation

$$-\Delta\rho - \mathbf{a} \cdot \nabla\rho = 0 \qquad \text{in the disk of radius } r > 0, \tag{7.4}$$

where (see again [3, eqns. (2.1–3) and (2.12)]), for $\mu \in (0, 2\pi)$,

$$0 \leq \omega := \frac{\|\mathbf{a}\|}{2}, \qquad \frac{\mathbf{a}}{2} = \omega(\cos\mu, \sin\mu) \in \mathbb{R}^2,$$

is the polar representation of the drift velocity $\mathbf{a} \in \mathbb{R}^2$ in (7.4). Accordingly, one can relate mesoscopic to macroscopic steady-states thanks to (7.2), and derivations similar to those performed in this article may lead to a "truly two-dimensional," asymptotic-preserving (in diffusive scaling) discretization of (7.1), in a similar manner as both (3.5) and (4.4).

**Appendix A. *S*-matrix for Goldstein–Taylor model in 1D.** It might be interesting to recall some properties of "two-stream" one-dimensional (position-dependent) radiative transfer, already studied in [16], [13, sect. 8.2], and [8, 11],

$$\partial_t f^{\pm} \pm \partial_x f^{\pm} = \sigma(x)(\rho/2 - f^{\pm}), \qquad \rho = f^+ + f^-.$$

Macroscopic (diffusive) stationary regimes in 1D reduce to $\rho''(x) = 0$, i.e., constant or linear functions, and yield Case's polynomial solutions, 1 and $x - v$. Accordingly, for $R = \Delta x/2$ and $f(x, v) = \alpha + \beta(x - v)$,

$$M = \begin{pmatrix} 1 & -(1+\sigma R) \\ 1 & (1+\sigma R) \end{pmatrix}, \qquad \tilde{M} = \begin{pmatrix} 1 & -(1-\sigma R) \\ 1 & (1-\sigma R) \end{pmatrix},$$

so that

$$|M| = 2(1+\sigma R), \qquad M^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -\frac{1}{1+\sigma R} & \frac{1}{1+\sigma R} \end{pmatrix},$$

meaning that $\alpha$ is the average of incoming states, and

$$S(\sigma) = \tilde{M}\, M^{-1} = \frac{1}{1+\sigma R}\begin{pmatrix} 1 & \sigma R \\ \sigma R & 1 \end{pmatrix}.$$

Such an $S$-matrix is "doubly stochastic" because both its rows and columns add to unity and all its entries are positive when $\sigma R \geq 0$. Asymptotic limits are

$$S(\sigma) \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ if } \sigma \to 0, \qquad S(\sigma) \to \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ if } \sigma \to +\infty.$$

The resulting well-balanced 1D time-marching scheme reads

$$\begin{pmatrix} f_j^{+,n+1} \\ f_{j-1}^{-,n+1} \end{pmatrix} = \left(1 - \frac{\Delta t}{2R}\right)\begin{pmatrix} f_j^{+,n} \\ f_{j-1}^{-,n} \end{pmatrix} + \frac{\Delta t}{2R}\, S(\sigma_{j-\frac{1}{2}})\begin{pmatrix} f_{j-1}^{+,n} \\ f_j^{-,n} \end{pmatrix}.$$

In parabolic scaling, the decomposition

$$S^{\varepsilon}(\sigma) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \frac{\varepsilon}{\varepsilon + \sigma R}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

holds and brings back the well-known IMEX scheme originally written in [16]:

$$\text{(A.1)} \qquad \begin{pmatrix} f_j^{+,n+1} \\ f_{j-1}^{-,n+1} \end{pmatrix}\left(1 + \frac{\Delta t}{2\varepsilon R}\right) - \frac{\Delta t}{2\varepsilon R}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} f_j^{+,n+1} \\ f_{j-1}^{-,n+1} \end{pmatrix}$$
$$= \begin{pmatrix} f_j^{+,n} \\ f_{j-1}^{-,n} \end{pmatrix} + \frac{\Delta t}{2R(\varepsilon + \sigma R)}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} f_{j-1}^{+,n} \\ f_j^{-,n} \end{pmatrix}.$$

Introducing the notation

$$A^n_{j-\frac{1}{2}} = \begin{pmatrix} f_j^{+,n} \\ f_{j-1}^{-,n} \end{pmatrix}, \qquad B^n_{j-\frac{1}{2}} = \begin{pmatrix} f_{j-1}^{+,n} \\ f_j^{-,n} \end{pmatrix},$$

one can even prove that the aforementioned IMEX scheme (A.1) is well balanced.

THEOREM A.1 (well-balanced IMEX). *Assuming* $2\Delta t \le \min(\sigma)R^2$ *and the data at time* $t^n$ *are at steady-state in the following sense, then they are invariant by* (A.1)*:*

$$\boxed{\forall i \in \mathbb{Z}, \qquad A^n_{j-\frac{1}{2}} = S^\varepsilon(\sigma_{j-\frac{1}{2}})\, B^n_{j-\frac{1}{2}} \quad \Longrightarrow \quad f_j^{\pm,n+1} = f_j^{\pm,n}.}$$

*Proof.* As $\varepsilon > 0$ is arbitrary, for simplicity, let $S = S^\varepsilon(\sigma_{j-\frac{1}{2}})$, which splits into

$$S = S_0 + \varepsilon\, S_1^\varepsilon, \qquad S_0 = \lim_{\sigma\to\infty} S(\sigma) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The IMEX scheme rewrites, for this particular type of data, as

$$\begin{aligned} A^{n+1}_{j-\frac{1}{2}} + \frac{\Delta t}{2\varepsilon R}\left(A^{n+1}_{j-\frac{1}{2}} - S_0\, B^{n+1}_{j-\frac{1}{2}}\right) &= A^n_{j-\frac{1}{2}} + \frac{\Delta t}{2\varepsilon R}\,\varepsilon S_1^\varepsilon\, B^n_{j-\frac{1}{2}} \\ &= A^n_{j-\frac{1}{2}} + \frac{\Delta t}{2\varepsilon R}(S - S_0)\, B^n_{j-\frac{1}{2}} \\ &= A^n_{j-\frac{1}{2}} + \frac{\Delta t}{2\varepsilon R}\left(A^n_{j-\frac{1}{2}} - S_0 B^n_{j-\frac{1}{2}}\right). \end{aligned}$$

By shifting indexes and taking advantage of $S_0$ not depending on $j$,

$$\begin{pmatrix} f_j^{+,n+1} \\ f_j^{-,n+1} \end{pmatrix} + \frac{\Delta t}{2\varepsilon R}\begin{pmatrix} f_j^{+,n+1} - f_j^{-,n+1} \\ f_j^{-,n+1} - f_j^{+,n+1} \end{pmatrix} = \begin{pmatrix} f_j^{+,n} \\ f_j^{-,n} \end{pmatrix} + \frac{\Delta t}{2\varepsilon R}\begin{pmatrix} f_j^{+,n} - f_j^{-,n} \\ f_j^{-,n} - f_j^{+,n} \end{pmatrix}.$$

Yet, adding both equations brings $\rho_j^{n+1} = \rho_j^n$, whereas subtracting them yields

$$\left(1 + \frac{\Delta t}{\varepsilon R}\right)\left(f_j^{+,n+1} - f_j^{-,n+1}\right) = \left(1 + \frac{\Delta t}{\varepsilon R}\right)\left(f_j^{+,n} - f_j^{-,n}\right),$$

and since $1 + \Delta t/\varepsilon R > 0$, this concludes the proof. ◻

## REFERENCES

[1] M. AINSWORTH AND W. DORFLER, *Fundamental systems of numerical schemes for linear convection-diffusion equations and their relationship to accuracy*, Computing, 66 (2001), pp. 199–229.

[2] G. BAL AND Y. MADAY, *Coupling of transport and diffusion models in linear transport theory*, Math. Model. Numer. Anal., 36 (2002), pp. 69–86.

[3] R. BIANCHINI AND L. GOSSE, *A truly two-dimensional discretization of drift-diffusion equations on Cartesian grids*, SIAM J. Numer. Anal., 56 (2018), pp. 2845–2870, https://doi.org/10.1137/17M1151353.

[4] G. BIRKHOFF AND I. ABU-SHUMAYS, *Harmonic solutions of transport equations*, J. Math. Anal. Appl., 28 (1969), pp. 211–221.

[5] A.V. BOBYLEV, *Exact solutions of discrete kinetic models and stationary problems for the plane Broadwell model*, Math. Methods Appl. Sci., 19 (1996), pp. 825–845.

[6] A.K. BOSE, *A mean value property of elliptic equations with constant coefficients*, Proc. Amer. Math. Soc., 18 (1967), pp. 995–996.

[7] T.A. Brunner and J.P. Holloway, *Two-dimensional time-dependent Riemann solvers for neutron transport*, J. Comput. Phys., 210 (2005), pp. 386–399.

[8] Ch. Buet, B. Despres, and T. Leroy, *Uniform convergence for a cell-centered AP discretization of the hyperbolic heat equation on general meshes*, Math. Comput., 86 (2017), pp. 1147–1202.

[9] Ch. Buet, B. Despres, and G. Morel, *Trefftz Discontinuous Galerkin Basis Functions for a Class of Friedrichs Systems Coming from Linear Transport*, https://hal.sorbonne-universite.fr/hal-01964528, 2018.

[10] J.G. Conlon, *Fundamental solutions for the anisotropic neutron transport equation*, Proc. Roy. Soc. Edinburgh Sect. A, 81 (1978), pp. 325–350.

[11] B. Despres and Ch. Buet, *The structure of well-balanced schemes for Friedrichs systems with linear relaxation*, Appl. Math. Comput., 272 (2016), pp. 440–459.

[12] L. Flatto, *Functions with a mean value property*, J. Math. Mech., 10 (1961), pp. 11–18.

[13] L. Gosse, *Computing Qualitatively Correct Approximations of Balance Laws: Exponential-Fit, Well-Balanced and Asymptotic-Preserving*, SIMAI Springer Ser. 2, Springer, Milan, 2013.

[14] L. Gosse, *Redheffer products and numerical approximation of currents in one-dimensional semiconductor kinetic models*, Multiscale Model. Simul., 12 (2014), pp. 1533–1560, https://doi.org/10.1137/130939584.

[15] L. Gosse, *A well-balanced and asymptotic-preserving scheme for the one-dimensional linear Dirac equation*, BIT, 55 (2015), pp. 433–458.

[16] L. Gosse and G. Toscani, *An asymptotic-preserving well-balanced scheme for the hyperbolic heat equations*, C. R. Math. Acad. Sci. Paris, 334 (2002), pp. 337–342.

[17] L. Gosse and N. Vauchelet, *Some examples of kinetic schemes whose diffusion limit is Il'in's exponential-fitting*, Numer. Math., 141 (2019), pp. 627–680, https://doi.org/10.1007/s00211-018-01020-8.

[18] A. Klar and N. Siedow, *Boundary layers and domain decomposition for radiative heat transfer and diffusion equations: Applications to glass manufacturing process*, European J. Appl. Math., 9 (1998), pp. 351–372.

[19] P.L. Lions and G. Toscani, *Diffusive limit for finite velocity Boltzmann kinetic models*, Riv. Mat. Iberoamericana, 13 (1997), pp. 473–513.

[20] C.-W. Shu, *Total-variation-diminishing time discretizations*, SIAM J. Sci. Stat. Comput., 9 (1988), pp. 1073–1084, https://doi.org/10.1137/0909073.

[21] O. Tretiak and C. Metz, *The exponential Radon transform*, SIAM J. Appl. Math., 39 (1980), pp. 341–354, https://doi.org/10.1137/0139029.

[22] X. Yang, F. Golse, Z. Huang, and S. Jin, *Numerical study of a domain decomposition method for a two-scale linear transport equation*, Networks Heterog. Media, 1 (2006), pp. 143–166.

[23] L. Zalcman, *Mean values and differential equations*, Israel J. Math., 14 (1973), pp. 339–352.