Conservative semi-Lagrangian schemes for Vlasov equations

Nicolas Crouseilles, Michel Mehrenberger, Eric Sonnendrücker

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Conservative semi-Lagrangian schemes for Vlasov equations

Nicolas Crouseilles — Michel Mehrenberger — Eric Sonnendrücker

N° 6856

February 2009

Thème NUM

ISSN 0249-6399 ISRN INRIA/RR--6856--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
NANCY - GRAND EST

Conservative semi-Lagrangian schemes for Vlasov equations

Nicolas Crouseilles*, Michel Mehrenberger†, Eric Sonnendrücker†

Thème NUM — Systèmes numériques
Équipe-Projet CALVI

Rapport de recherche n° 6856 — February 2009 — 32 pages

Abstract: Conservative methods for the numerical solution of the Vlasov equation are developed in the context of the one-dimensional splitting. In the case of constant advection, these methods and the traditional semi-Lagrangian ones are proven to be equivalent, but the conservative methods offer the possibility to add adequate filters in order to ensure the positivity. In the non constant advection case, they present an alternative to the traditional semi-Lagrangian schemes which can suffer from bad mass conservation, in this time splitting setting.

Key-words: Vlasov equation, semi-Lagrangian method, finite volume methods, conservative methods, plasma physics

* INRIA Nancy Grand Est and IRMA (Université de Strasbourg and CNRS)
† IRMA (Université de Strasbourg and CNRS) and INRIA Nancy Grand Est

Centre de recherche INRIA Nancy – Grand Est
LORIA, Technopôle de Nancy-Brabois, Campus scientifique,
615, rue du Jardin Botanique, BP 101, 54602 Villers-Lès-Nancy
Téléphone : +33 3 83 59 30 00 — Télécopie : +33 3 83 27 83 19

Méthodes semi-Lagrangiennes conservatives pour la résolution numérique de l'équation de Vlasov

Résumé : Des méthodes conservatives sont dveloppées pour la résolution numérique de l'équation de Vlasov dans le cadre de splitting directionnel unidimensionnel. Dans le cas de l'advection à coefficient constant, ces méthodes sont quivalentes aux méthodes semi-Lagrangiennes traditionnelles, mais pour les méthodes conservatives, il est possible d'ajouter des filtres qui garantissent la positivité de la solution. Dans le cas non-constant, les méthodes conservatives sont une alternative aux méthodes semi-Lagrangiennes classiques. Ces dernières peuvent prsenter des mauvaises conservations de la masse totale lorsqu'une technique de splitting est utilisée.

Mots-clés : Schéma semi-Lagrangien, méthodes des volumes finis, méthodes conservatives, simulation numérique pour Vlasov, physique des plasmas

1 Introduction

To describe the dynamics of charged particles in a plasma or in a propagating beam, the Vlasov equation can be used to calculate the plasma response to the electromagnetic fields. The unknown $f(t, x, v)$ which depends on the time t, the space x and the velocity v represents the distribution function of the studied particles. The coupling with the self-consistent electromagnetic fields is taken into account through the Maxwell or Poisson equation.

Due to its nonlinear structure, analytical solutions are available only in few academic cases, and numerical simulations have to be performed to study realistic physical phenomena. Nowadays, mostly two classes of methods are used to investigate the behaviour of the numerical solution to the Vlasov equation. On the one hand, Particle In Cell (PIC) methods, which are the most widely used, approach the plasma by macro-particles, the trajectories of which follow the characteristic curves of the Vlasov equation whereas the electromagnetic fields are computed by gathering the charge and current densities particles on a grid of the physical space (see [2]). On the other hand, Eulerian methods consist in discretizing the Vlasov equation on a grid of the phase space using classical numerical schemes such as finite volumes or finite elements methods for example (see [6, 9, 22]).

Although PIC methods can theoretically and potentially resolve the whole 6 dimensional problem, it is well known that the inherent numerical noise makes difficult a precise description of low density regions, despite significant recent improvements. Hence, Eulerian methods offer a good alternative to overcome this lack of precision, even if problems of memory can arise when one deals with high dimensions. In particular, Vlasov codes seem to be appropriate to study nonlinear processes.

This last decade, gridded Vlasov solvers have been developed for $2D$, $4D$ and even $5D$ phase space problems. Among them, the semi-Lagrangian method using a cubic spline interpolation (SPL) [22] and the Positive Flux Conservative (PFC) method [9] have been implemented to deal with physical applications [12, 11, 24].

Recently, a parabolic spline method (PSM) has been introduced for transport equations arising in meteorology applications [26, 27]. This mehod benefits from the best approximation property of the SPL method and from the conservation of mass and positivity (by applying a suitable filter) of the PFC method.

The aim of the present work is to study such a conservative method in the context of the Vlasov equation. Conservative methods present a lot of advantages. In addition to the inherent conservative property, slope limiters can be introduced in the reconstruction to ensure positivity; moreover, since they solve the conservative form of the equation, multi-dimensional problems can be solved by a splitting procedure so that the solution of the full problem is reduced to a succession of solution to only one-dimensional problems. Obviously, this property is of great interest from an implementation and algorithmic point of view.

We will focus here essentially on PSM, which has never been applied to our knowledge to Vlasov simulations. We will also introduce a new method based on a cubic splines approximation of the unknown; the characteristics curves are followed forwardly as in [7], but the unknown is reconstructed in a conservative way using its values on the transported non-uniform mesh.

In our numerical experiments, we first consider the special case of directional splitting with constant advection (like the Vlasov-Poisson system). In that case, when no filter is applied, we prove that the advective scheme (e.g. SPL) and the conservative one (e.g. PSM) are *equivalent*. Note that in this setting, a mathematical proof of the convergence has been performed in [1]. We then discuss the choice of the filter in order to preserve the positivity. The filters introduced in [26, 27] seem to modify too much the distribution function; we thus propose new filters which minimize in a certain sense the gap with the initial reconstruction while maintaining the positivity constraint. Numerical results are given for a classical plasma test case: the bump-on-tail instability.

We then focus on the case where we do not have a constant advection, as is the case for the guiding center model. In [22], a $2D$ interpolation was proposed to approximate this model to ensure the conservation property (which is not true if a splitting procedure is used [14]). The time splitting in the conservative form has also been successfully tested for the PFC scheme [3], which is however more diffusive. Different works have been devoted to the study of the CIP method in its conservative form (see [18, 19, 23] and references therein). The time splitting in the advective form is often discarded since it can lead to bad mass conservation, especially for long time simulations ([14, 18]). We see here that with the conservative spline formulation (PSM), we can perform simulations with directional splitting not as diffusive as PFC, while maintaining the mass conservation. The time step is nevertheless limited by a CFL condition which seems luckily not to be so severe in our context. In particular, we obtain good numerical results for the classical test cases occuring in plasma physics.

The paper is organized as follows: first, semi-Lagrangian conservative methods are recalled and also introduced for one-dimensional general problems. Then, the constant advection case is investigated, and it is proved that, in this case, a conservative method and its advective counterparts are equivalent when no filter are considered; numerical results applied to the Vlasov-Poisson model are then discussed. Finally, we focus on the more interesting non-constant advection case for which numerical results illustrate the good behaviour of the new approaches.

2 Conservative methods

We are interested in the approximation of multi-dimensional transport equations of the form

$$\frac{\partial g}{\partial t} + \nabla_x \cdot (ag) = 0, \quad x \in \Omega \subset \mathbb{R}^n, \tag{1}$$

where the unknown g depends on time and on the multi-dimensional spatial direction x and a is a divergence free vector field $\nabla_x \cdot a = 0$ which can depend on time. The so-called conservative form (1) is then equivalent to the advective form

$$\frac{\partial g}{\partial t} + a \cdot \nabla_x g = 0, \quad x \in \Omega \subset \mathbb{R}^n. \tag{2}$$

In Vlasov type equations which enter in the class of equation of the form (1), Ω is a subset of the phase space which has up to 6 dimensions.

Splitting the components of x into x_1 and x_2, equation (1) can be written in the form

$$\frac{\partial g}{\partial t} + \nabla_{x_1} \cdot (a_1 g) + \nabla_{x_2} \cdot (a_2 g) = 0,$$

where a_1 and a_2 denote the component of the field a corresponding to x_1 and x_2. It is well known (see [6]) that a splitting procedure involves a successive solution of

$$\frac{\partial g}{\partial t} + \nabla_{x_1} \cdot (a_1 g) = 0, \quad \frac{\partial g}{\partial t} + \nabla_{x_2} \cdot (a_2 g) = 0, \tag{3}$$

keeping high order accuracy in time for the whole equation (1). However, the traditional semi-Lagrangian methods described in [22] for example do not resolve the conservative form but the non-conservative form of the equations (2). Then, by solving only the advective form of (3), the corrective terms $g\nabla_{x_1} \cdot a_1$ and $g\nabla_{x_2} \cdot a_2$ are omitted and can lead to an important lack of accuracy in long time simulations (see [14, 18]). An alternative way would be to solve the conservative form so that the solution of (1) can be performed by solving a succession of one-dimensional problems. Hence in the sequel, we propose different conservative methods to solve one-dimensional problems; the methods are first presented in a general context but practical examples will be detailed in the next sections.

2.1 Conservative semi-Lagrangian methods for one-dimensional problems

The conservative methods enable to solve the conservative terms separately so that we restrict ourselves to the following one-dimensional problem

$$\frac{\partial g}{\partial t} + \frac{\partial (ag)}{\partial x} = 0, \quad x \in I \subset \mathbb{R}. \tag{4}$$

Let us consider a grid of the interval I: $x_{-1/2} < x_0 < x_{1/2} < x_1 < ... < x_{N-1} < x_{N-1/2} < x_N$. We consider the average quantity

$$\bar{g}_i^n = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} g(t^n, x) dx, \quad \text{with} \quad \Delta x = x_{i+1/2} - x_{i-1/2}, \quad i = 0, .., N-1.$$

Thanks to the conservation of the volume, we can write the following equality

$$\int_{x_{i-1/2}^{n+1}}^{x_{i+1/2}^{n+1}} g(t^{n+1}, x) dx = \int_{x_{i-1/2}^{n}}^{x_{i+1/2}^{n}} g(t^n, x) dx, \tag{5}$$

where $x_{i-1/2}^{n+1}$ and $x_{i-1/2}^n$ belong to the same characteristic curve defined by

$$\frac{dX(t)}{dt} = a(t, X(t)), \quad X(t^{n+1}) = x_{i-1/2}^{n+1}, \; X(t^n) = x_{i-1/2}^n, \quad i = 0, .., N-1.$$

Assuming the values of g are known at time t^n on each volume $[x_{i-1/2}, x_{i+1/2}]$, we can reconstruct the primitive function $G(t^n, x)$ of $g(t^n)$ in I. Let us remark that the primitive function is only known on the mesh. To do this, we define $G(t^n)$ as a cumulative function corresponding to $\bar{g}^n$:

$$G(t^n, x_{i-1/2}) = \Delta x \sum_{k=0}^{i-1} \bar{g}_k^n, \quad i = 0, .., N. \tag{6}$$

Hence, using (5), we have

$$\begin{aligned}\bar{g}_i^{n+1} &= \int_{x_{i-1/2}^{n+1}}^{x_{i+1/2}^{n+1}} g(t^{n+1},x)dx \\ &= \int_{x_{i-1/2}^{n}}^{x_{i+1/2}^{n}} g(t^{n},x)dx \\ &= G(t^n,x_{i+1/2}^n)-G(t^n,x_{i-1/2}^n).\end{aligned}$$

The only thing to do is to build a polynomial primitive function, assuming $G(t^n)$ is known on the mesh $(x_{i-1/2})_{i=0,..,N-1}$. Several ways can be used to achieve a good approximation.

Hence, as in the pointwize semi-Lagrangian method, the algorithm of conservative methods is composed of two main steps: the computation of the characteristic curves, and the reconstruction step.

2.1.1 Computation of the characteristic curves

In the semi-Lagrangian method, we have to compute the characteristics curves between two consecutive time steps. As remarked in [22], second order accuracy can be reached using a two steps approach. Indeed, if we assume that g^{n-1} and g^n are known, we can compute the advection term $a(t^n,X(t^n))$ which can depend on the solution g^n. Let us remark that the advection term is assumed constant in time between t^{n-1} and t^{n+1}. Then, the equation to solve writes

$$\frac{dX(t)}{dt}=a(t^n,X(t)), \tag{7}$$

with a final condition which coincides with the mesh point X_i

$$X(t^{n+1})=X_i=x_0+i\Delta x. \tag{8}$$

Here, (7) is solved using a parabolic assumption (see [22, 12]). Let X_i be the position of $X(t^{n+1})$, then there exists a displacement $\alpha_i\in\mathbb{R}$ such that

$$X_i^n=X_i-\alpha_i, \quad \text{and} \quad X_i^{n-1}=X_i-2\alpha_i.$$

Hence, the displacement can be computed at second order by solving the following one-dimensional fixed-point

$$\alpha_i=\Delta t\ a(t^n,X_i-\alpha_i). \tag{9}$$

In [22], a Newton algorithm is used but every iterative methods can be employed. We also mention [12] in which a Taylor expansion of the right hand side of (9) is performed; this strategy is equivalent to a Newton algorithm in which two iterations are imposed. However, the drawback of these algorithms is that they require the evaluation of the Jacobian matrix of $a(t^n)$. A fixed point algorithm can then be implemented. But, if we assume linear reconstruction of the advection term at points $(X_i-\alpha_i)$ (as it is supposed in [22, 12]), an explicit algorithm can be used. The main steps of this new algorithm is detailed in the sequel.

Starting from (9) and denoting by $[x_j, x_{j+1}]$ the cell in which $X_i - \alpha_i$ falls, the linear reconstruction of $a(t^n)$ writes

$$\alpha_i = \Delta t \ [(1-\beta)a(t^n, x_j) + \beta a(t^n, x_{j+1})], \tag{10}$$

where β is such that

$$X_i - \alpha_i = x_j + \beta, \ \ x_j = x_0 + j\Delta x, \ \ X_i = x_0 + i\Delta x. \tag{11}$$

Injecting the expression of α_i into (10) leads

$$\beta\left[\Delta x + \Delta t\left(a(t^n, x_{j+1}) - a(t^n, x_j)\right)\right] = (i-j)\Delta x - \Delta t\, a(t^n, x_j), \tag{12}$$

from which an expression of β can be deduced

$$\beta = \left[(i-j)\Delta x - \Delta t\, a(t^n, x_j)\right] / \left[\Delta x + \Delta t(a(t^n, x_{j+1}) - a(t^n, x_j))\right]. \tag{13}$$

Now, it remains to determine the j index. To do that, it must be remarked that β given by (13) lives in the interval $[0, 1]$. Hence, from (12), we can deduce an expression for $x_i = i\Delta x$

$$i\Delta x = j\Delta x + \Delta t \ a(t^n, x_j) + \beta\left[\Delta x + \Delta t(a(t^n, x_{j+1}) - a(t^n, x_j))\right].$$

Using the fact that $\beta \in [0, 1]$, and by remarking that $[\Delta x + \Delta t(a(t^n, x_{j+1}) - a(t^n, x_j))] > 0$ provided that Δt is small enough, we deduce

$$i\Delta x \in [M_j, M_{j+1}], \ \ \text{with } M_j = x_j + \Delta t\, a(t^n, x_j).$$

Under the assumption that Δt is small enough, the non-decreasing sequence $(M_j)_{j=0,.,N-1}$ forms a non-uniform mesh from which it can be easily found the location of X_i. The algorithm is then the following for each $i = 0, .., N-1$:

- determination of j such that $X_i \in [M_j, M_{j+1}]$
- determination of β with (13)
- determination of α_i with (11)

This algorithm has been proved to be faster than classical iterative based methods. Obviously, it leads to the same displacement α_i.

2.1.2 Formulation by primitive

In this subsection, we present some reconstructions to update the unknown; once the feet of the characteristics is computed, we have to evaluate the integral of the function $g(t^n)$. Two approaches to do that are presented in the sequel.

Lagrange reconstruction of the primitive. This paragraph presents a Lagrange reconstruction of the primitive, as done in [9] for example. The primitive is known on the mesh $x_{i-1/2}, i = 1, ..., N$, so that the cubic Lagrange polynomial can be written

$$G(t^n, x) = \sum_{j=i-1}^{i+2} G(t^n, x_{j-1/2}) L_j(x), \ \ \forall x \in [x_{i-1/2}, x_{i+1/2}], \tag{14}$$

where

$$L_j(x) = \prod_{k=j-1, k\neq j}^{j+2} (x - x_{k-1/2})/(x_{j-1/2} - x_{k-1/2}).$$

This reconstruction corresponds to the PFC method introduced in [9] in which the slope limiters step is not performed. This approach and similar ones has been also introduced by [16, 5]. See [17] for a more complete bibliography.

PSM: Cubic splines reconstruction of the primitive. We can also build a cubic spline for the interpolation conditions of which are given by (6). This can be written as

$$G(t^n, x) = \sum_{i=0}^{N-1} \eta_i S\left(\frac{x - x_{i-1/2}}{\Delta x}\right), \tag{15}$$

where η_i are the coefficients and S the cubic splines

$$6S(x) = \begin{cases} (2-|x|)^3 & \text{if } 1 \leq |x| \leq 2, \\ 4 - 6x^2 + 3|x|^3 & \text{if } 0 \leq |x| \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

In addition to the interpolation conditions, we close the system satisfied by the cubic splines coefficients η_i by considering the periodic case. But, even if $\bar{g}^n$ is periodic ($\bar{g}^n_0 = \bar{g}^n_N$), let us mention that it is not true for the cumulative function. However, we have the following relation on the primitive for $x > x_{N-1/2}$

$$\begin{aligned} G(t^n, x) &= \int_{x_{-1/2}}^{x} g(t^n, x)dx = \int_{x_{-1/2}}^{x_{N-1/2}} g(t^n, x)dx + \int_{x_{N-1/2}}^{x} g(t^n, x)dx \\ &= M + \int_{x_{N-1/2}}^{x} g(t^n, x)dx = M + G(t^n, x), \end{aligned}$$

where M denotes the total mass $\sum_{k=0}^{N-1} \bar{g}^n_k$. Then, using this last property, a periodic linear system can be recovered

$$A\eta = \begin{pmatrix} 4 & 1 & 0 & 0 & \cdots & 1 \\ 1 & 4 & 1 & 0 & & \vdots \\ 0 & 1 & 4 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & 0 & 1 & 4 & 1 \\ 1 & 0 & 0 & 0 & 1 & 4 \end{pmatrix} \begin{pmatrix} \eta_0 \\ \eta_1 \\ \vdots \\ \vdots \\ \eta_{N-2} \\ \eta_{N-1} \end{pmatrix} = 6\Delta x \begin{pmatrix} \bar{g}_0 + M \\ \bar{g}_0 + \bar{g}_1 \\ \vdots \\ \vdots \\ \sum_{i=0}^{N-2} \bar{g}_i \\ \sum_{i=0}^{N-1} \bar{g}_i - M \end{pmatrix}$$

The coefficients $\eta_i, i \notin [0, N-1]$ are deduced from the solutions of the previous linear system by

$$\eta_{-i} = \eta_{-i+N} - M, \ \ \forall i > 0, \quad \eta_{i+N-1} = \eta_i + M, \ \ \forall i \in [0, N-1].$$

This approach has been introduced in [27] as the Parabolic Spline Method. Their formulation refers to the reconstruction of the function g which is a $\mathcal{C}^1$ piecewise parabolic function. The two formulations (by using primitive G or the function g) are completely equivalent, as already explained in [27].

2.2 Forward update of the characteristics

Another strategy to update in a conservative way the unknowns consists in advancing in time the mesh points which are denoted by $x^n_{i-1/2}, i = 0, .., N$. The ends of the characteristics starting at the mesh $x^n_{i-1/2}$ are called $x^{n+1}_{i-1/2}$, $i = 0, .., N$. These points form a non-uniform mesh of the domain in the general case. Then, due to the property of the conservation of the volumes (5), it is possible to reconstruct a primitive function $G(t^{n+1})$ of $g(t^{n+1})$ with cubic spline polynomials in the spirit of paragraph 2.1.2. The main differences of the present case are twofold. On the one side, the mesh on which $G(t^{n+1})$ is known and has to be reconstructed is not uniform, and on the other side, the characteristic curves are advanced in time.

In the rest of the section, the two steps of the method are detailed.

2.2.1 Computation of the characteristics curves

We have to compute the characteristics curves between two consecutive time steps. As previously, if we assume that g^{n-1} and g^n are known, we can compute the advection term $a(t^n, X(t^n))$ which can depend on the solution g^n. Let us remark that the advection term is supposed constant in time between t^{n-1} and t^{n+1} so that the equation to solve writes

$$\frac{dX(t)}{dt} = a(t^n, X(t)), \tag{16}$$

with an initial condition which coicides with a mesh point X_i in the considered case

$$X(t^{n-1}) = X_i = x_0 + i\Delta x. \tag{17}$$

Even if the forward non-uniform approach can be expensive from a numerical point of view due to the solution of a new linear system at each time step, this algorithm has the opportunity to compute the characteristics curves in an explicit way. Indeed, contrary to the traditional backward method, the initial condition of (16) is given at the initial time. Hence, explicit algorithms can be implemented like Runge-Kutta ones. In our practical experiments, methods up to the fourth order Runge-Kutta algorithm have been implemented even if the formulation (9) is only second order accurate by essence. Let us recall the Runge-Kutta 2 method applied to our problem (we suppose that the advection field remains constant in time during the following steps, equal to its value at time t^n)

$$k_1 = \mathcal{I}a(t^n, X(t^{n-1})), \quad k_2 = \mathcal{I}a(t^n, X(t^{n-1}) + 2\Delta t\, k_1),$$

which leads to the following approximation of the characteristics at time t^{n+1}

$$X(t^{n+1}) = X(t^n) + \frac{2\Delta t}{2}(k_1 + k_2).$$

We denote by $\mathcal{I}$ an interpolation operator. In our experiments, a cubic spline polynomial has been used. Note that in the context of the time splitting studied in the guiding center model, we have also successfully applied Runge Kutta schemes for the backward method.

2.2.2 Interpolation using a non-uniform mesh

The so-reconstructed primitive G can then be interpolated on the uniform mesh $x_{i-1/2}, i = 0, .., N$ to obtain the update unknown in the following way

$$G(t^{n+1}, x_{i+1/2}) - G(t^{n+1}, x_{i-1/2}) = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} g^{n+1}(x)dx = \bar{g}_i^{n+1}.$$

As we said, the main step consists in the computation of the cubic splines coefficients on a non-uniform and periodic mesh. Cubic splines relations can be obtained by a repetition of convolution of the zero-th order splines (the box function) with itself (see [8]). Restricting to third order, we are faced to the following tridiagonal system to solve

$$A\eta = \begin{pmatrix} D_0 & C_1 & 0 & 0 & \cdots & B_{N-1} \\ B_0 & D_1 & C_2 & 0 & & \vdots \\ 0 & B_1 & D_2 & C_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & 0 & B_{N-3} & D_{N-2} & C_{N-2} \\ C_1 & 0 & 0 & 0 & B_{N-2} & D_{N-1} \end{pmatrix} \begin{pmatrix} \eta_0 \\ \eta_1 \\ \vdots \\ \vdots \\ \eta_{N-2} \\ \eta_{N-1} \end{pmatrix} = \begin{pmatrix} \bar{g}_0 + B_{N-1}M \\ \bar{g}_0 + \bar{g}_1 \\ \vdots \\ \vdots \\ \sum_{i=0}^{N-2} \bar{g}_i \\ \sum_{i=0}^{N-1} \bar{g}_i - C_0 M \end{pmatrix}$$

where the components of the matrix are defined for $i = 0, .., N-1$

$$B_i = \Delta x_{i+1}^2 / [(\Delta x_{i+1} + \Delta x_i + \Delta x_{i-1})(\Delta x_{i+1} + \Delta x_i)]$$

$$\begin{aligned} D_i &= (\Delta x_{i-2} + \Delta x_{i-1})\Delta x_i / [(\Delta x_i + \Delta x_{i-1} + \Delta x_{i-2})(\Delta x_i + \Delta x_{i-1})] \\ &+ (\Delta x_{i+1} + \Delta x_i)\Delta x_i / [(\Delta x_{i+1} + \Delta x_i + \Delta x_{i-1})(\Delta x_i + \Delta x_{i-1})] \end{aligned}$$

$$C_i = \Delta x_{i-2}^2 / [(\Delta x_i + \Delta x_{i-1} + \Delta x_{i-2})(\Delta x_{i-1} + \Delta x_{i-2})],$$

and Δx_i is defined as follows: $\Delta x_i = x_{i+1/2}^{n+1} - x_{i-1/2}^{n+1}$, $i = 0, .., N-1$ and defined for all i by periodicity. As in subsection 2.1.2, the periodicity can be taken into account in the linear system, even if the primitive is not periodic.

Let us mention that a good behaviour of the present method requires a good approximation of the characteristics curves $x_{i+1/2}^{n+1}$ at time t^{n+1}. Indeed, this strong dependence can be explained by the fact that they refer to the conditions of interpolation.

2.3 Slope limiters

In this subsection, we focus on the description of different filters which can be adapted to the previous reconstruction. Various approches have already been introduced (see [9, 13, 27]). We discuss here the use of specific filters for the PFC and PSM methods.

The PFC method In order to ensure the positivity, we are faced to the following problem: given $\bar{g}_{i-1}, \bar{g}_i, \bar{g}_{i+1}$, we have to find a reconstruction $P = P(\bar{g}_{i-1}, \bar{g}_i, \bar{g}_{i+1})$ which satisfies

$$\frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} P(x)\mathrm{d}x = \bar{g}_i, \quad P(x) \geq 0, \quad x \in (x_{i-1/2}, x_{i+1/2}). \tag{18}$$

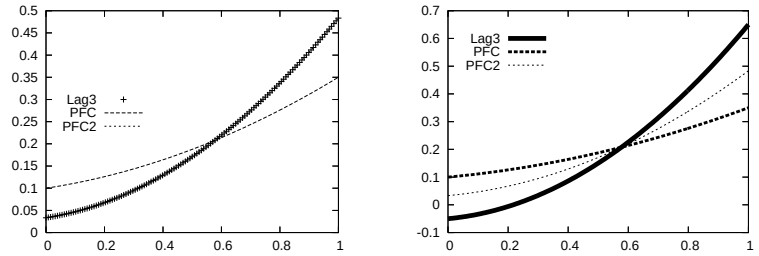

Figure 1: Comparison of the semi-Lagrangian LAG3, PFC and PFC2 reconsructions.

The reconstruction on $(x_{i-1/2}, x_{i+1/2})$ with the PFC method without the filter is the unique polynomial of degree ≤ 2 that satisfies

$$\frac{1}{\Delta x}\int_{x_{i-3/2}}^{x_{i-1/2}} P(x)\mathrm{d}x = \bar{g}_{i-1}, \quad \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} P(x)\mathrm{d}x = \bar{g}_{i-1}, \quad \frac{1}{\Delta x}\int_{x_{i+1/2}}^{x_{i+3/2}} P(x)\mathrm{d}x = \bar{g}_{i+1}.$$

By explicit computations, we have

$$P(x_{i-1/2}+\omega\Delta x) = \bar{g}_i-(\bar{g}_i-\bar{g}_{i-1})((1-\omega)^2-1/3)/2+(\bar{g}_{i+1}-\bar{g}_i)(\omega^2-1/3)/2, \quad 0\leq\omega\leq 1.$$

We define $\bar{g}_{\min} = \min_{i=0,\ldots,N-1} \bar{g}_i$, $\bar{g}_{\max} = \max_{i=0,\ldots,N-1} \bar{g}_i$ and suppose that $\bar{g}_{\min} \geq 0$, which should be the case for our initial data.
We propose here to take the following filter, which is a little less restrictive version of the one used in [9]

$$P(x_{i-1/2}+\omega\Delta x) = \bar{g}_i-\epsilon^-(\bar{g}_i-\bar{g}_{i-1})((1-\omega)^2-1/3)/2+\epsilon^+(\bar{g}_{i+1}-\bar{g}_i)(\omega^2-1/3)/2, \quad 0\leq\omega\leq 1,$$

with

$$(\epsilon^-,\epsilon^+) = \begin{cases} \big(\min(1,4\dfrac{\bar{g}_{\max}-\bar{g}_i}{\bar{g}_i-\bar{g}_{i-1}}),\min(1,4\dfrac{\bar{g}_i-\bar{g}_{\min}}{\bar{g}_{i+1}-\bar{g}_i})\big), & \text{if } \bar{g}_i>\bar{g}_{i-1} \text{ and } \bar{g}_{i+1}>\bar{g}_i, \\ \big(\min(1,3\dfrac{\bar{g}_{\max}-\bar{g}_i}{\bar{g}_i-\bar{g}_{i-1}}),\min(1,3\dfrac{\bar{g}_{\max}-\bar{g}_i}{\bar{g}_i-\bar{g}_{i+1}})\big), & \text{if } \bar{g}_i>\bar{g}_{i-1} \text{ and } \bar{g}_{i+1}<\bar{g}_i, \\ \big(\min(1,4\dfrac{\bar{g}_i-\bar{g}_{\min}}{\bar{g}_{i-1}-\bar{g}_i}),\min(1,4\dfrac{\bar{g}_{\max}-\bar{g}_i}{\bar{g}_i-\bar{g}_{i+1}})\big), & \text{if } \bar{g}_i<\bar{g}_{i-1} \text{ and } \bar{g}_{i+1}<\bar{g}_i, \\ \big(\min(1,3\dfrac{\bar{g}_i-\bar{g}_{\min}}{\bar{g}_{i-1}-\bar{g}_i}),\min(1,3\dfrac{\bar{g}_i-\bar{g}_{\min}}{\bar{g}_{i+1}-\bar{g}_i})\big), & \text{if } \bar{g}_i<\bar{g}_{i-1} \text{ and } \bar{g}_{i+1}>\bar{g}_i. \end{cases} \tag{19}$$

The reconstruction then satisfies (18) and even $\bar{g}_{\min} \leq P(x) \leq \bar{g}_{\max}$, for $x \in (x_{i-1/2}, x_{i+1/2})$. We will refer to this method as the PFC2 method (see Figs. 1 for two examples of positive reconstructions and the corresponding polynomial Lagrange 3 (LAG3) reconstruction).

The PSM method We denote by P_i the reconstruction of PSM without the filter on the interval $(x_{i-1/2}, x_{i+1/2})$. Explicit computations give for $0<\omega<1$,

$$P_i(x_{i-1/2}+\omega\Delta x) = g_{i-1/2}+\omega(g_{i+1/2}-g_{i-1/2})+(6\bar{g}_i-3(g_{i-1/2}+g_{i+1/2}))\omega(1-\omega),$$

where we have set $g_{i+1/2} = P_i(x_{i+1/2}) = P_{i+1}(x_{i+1/2})$, for $i = 0, \dots, N-1$. In order to satisfy the positivity for $P_i(x)$, we proceed in two steps, like in [26, 27], but we want to relax the filter used there which may be too severe in our context.

1. We take $g^{\rm new}_{i+1/2} = \max(\bar{g}_{\min}, \min(\bar{g}_{\max}, g_{i+1/2}))$, so that we have $\bar{g}_{\min} \leq g^{\rm new}_{i+1/2} \leq \bar{g}_{\max}$, $\forall i$. We then consider the first reconstruction for $\omega \in [0,1]$

$$R_1(x_{i-1/2}+\omega\Delta x) = g^{\rm new}_{i-1/2}+\omega(g^{\rm new}_{i+1/2}-g^{\rm new}_{i-1/2})+(6\bar{g}_i-3(g^{\rm new}_{i-1/2}+g^{\rm new}_{i+1/2}))\omega(1-\omega),$$

2. If $R_1(x_{i-1/2}+\omega\Delta x)$ has a strictly negative value for some $\omega_0 \in (0,1)$, we consider for $0 \leq \beta \leq 1$ the unique polynomial $R_{2,\beta}$ of degree ≤ 2 which satisfies

$$\int_{x_{i-1/2}}^{x_{i+1/2}} R_{2,\beta}(x){\rm d}x = \bar{g}_i,\ R_{2,\beta}(x_{i-1/2}+\beta\Delta x) = 0,\ R'_{2,\beta}(x_{i-1/2}+\beta\Delta x) = 0. \tag{20}$$

Clearly, this polynomial remains positive for $x \in (x_{i-1/2}, x_{i+1/2})$. Now, among all $\beta \in [0,1]$ which satisfy (20), we want to choose the closest polynomial $R_{2,\beta}$ to R_1 in a certain sense. We choose here β^* that minimizes a functional $J(\beta)$. This functional can be chosen according to a property of interest which we want to respect as best as possible. For example, if we want to reconstruct a polynomial function $R_{2,\beta}$ defined as in (20), the boundary cell values of which are as close as possible to $g^{\rm new}_{i\pm1/2}$, we consider the following functional

$$J(\beta) = |g^{\rm new}_{i-1/2} - R_{2,\beta}(x_{i-1/2})| + |g^{\rm new}_{i+1/2} - R_{2,\beta}(x_{i+1/2})|. \tag{21}$$

The minimum of this functional is identified at some $\beta^\star$ and thus we take $R = R_{2,\beta^*}$. In the case where $R_1(x_{i-1/2}+\omega\Delta x)$ has no strictly negative value, for $\omega_0 \in (0,1)$, we do no other change, that is, we take $R = R_1$. Hence the whole reconstruction satisfies the positivity of the polynomial reconstruction (see Figs. 2 in which for we compare the filter of [26] (PSM ZER), our optimal filter and the corresponding cubic splines).

Note that the approach is general and we can use other functionals. We can for example optimize the following functional

$$J(\beta) = \left| \int_{x_{i-1/2}}^{x_{i+1/2}} x^n \left(R_1(x) - R_{2,\beta}(x)\right) dx \right|, n = 1, 2.$$

The functional reaches its minimum at $\beta = \beta^\star$. When $n = 1$, the $R_{2,\beta^\star}$ reconstruction is as close as possible to the R_1 reconstruction regarding the first momentum which is preserved by the R_1 reconstruction. It is also possible to consider higher moments through the functional ($n = 2$ for example), but in these cases, the first momentum is not well preserved in our numerical tests, which pollutes the others diagnostics.

More generally, it is possible to dissociate the first reconstruction R_1 (which can be done with a Lagrange polynomial as for the PFC method) from the second one. We could also minimize in the L^2 norm (see e.g. [13] for an analog of (21) with squares instead of absolute values). Generally, it is simpler to deal with squares than with absolute values for the search of a minimum. In that case, it turns out that it is the opposite case : for the squares, we have to search the roots of polynomials of degree 4 and only of degree 2 for absolute values.

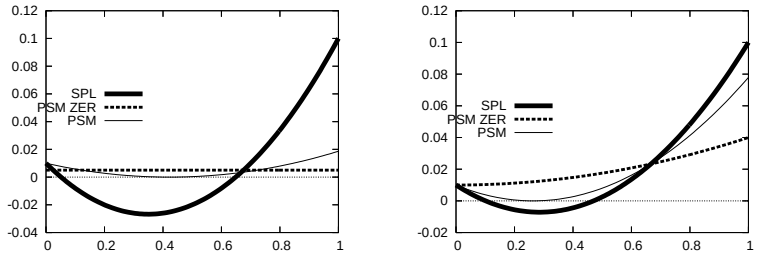

Figure 2: Comparison of the semi-Lagrangian cubic splines SPL, PSM ZER of [26] and PSM reconstructions.

3 The constant advection case

In this section, we are concerned with a constant advection field, so (4) can be rewritten equivalently in a conservative form

$$\frac{\partial g}{\partial t} + \frac{\partial (ag)}{\partial x} = 0, \quad x \in I \subset \mathbb{R},$$

or in an advective form

$$\frac{\partial g}{\partial t} + a\frac{\partial g}{\partial x} = 0, \quad x \in I \subset \mathbb{R}.$$

Let us define for $\alpha \in \mathbb{R}$ and $N \in \mathbb{N}^*$ a transport operator $\mathcal{T}_{\alpha,N} : \mathbb{R}^N \to \mathbb{R}^N$. We will define several possibilities for the transport operator. If $(g_0, \ldots, g_{N-1})$ is a discretization of a function g, then $\mathcal{T}_{\alpha,N}(g_0, \ldots, g_{N-1})$ should be a discretization of the shifted function $x \to g(x + \alpha)$. In our case α denotes the displacement given by the solution of (7) in which a is considered constant (in time and in space); we deduce that $\alpha = -\Delta t\, a$.

3.1 The advective approach

We reconstruct a function $\tilde{g}^a$ which satisfies

$$\tilde{g}^a(x_i) = g_i, \ i = 0, \ldots, N-1, \tag{1}$$

and we take

$$\mathcal{T}_{\alpha,N}(g_0, \ldots, g_{N-1})_i = \tilde{g}^a(x_i + \alpha). \tag{2}$$

We suppose periodic boundary conditions, that is $g_i = g_{i+kN}$ for $k \in \mathbb{Z}$. We give here some standard interpolating functions $\tilde{g}$ (see e.g. [10, 22]).

Lagrange interpolation of order $d = 2p + 1$ ($\mathbf{Lag_d}$) : if $x \in (x_k, x_{k+1})$, we take $\tilde{g}^a(x) = P(x)$, where P is the unique interpolating polynomial of degree $\leq d$ such that

$$P(x_i) = g_i, \text{ for } i = k - p, \ldots, k + p + 1.$$

Cubic spline interpolation (SPL) : we define the cubic spline S by

$$6S(x) = \begin{cases} (2 - |x|)^3 & \text{if } 1 \leq |x| \leq 2, \\ 4 - 6x^2 + 3|x|^3 & \text{if } 0 \leq |x| \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

and take $\tilde{g}^a(x) = \sum_{i=0}^{N-1} \eta_i S(\frac{x - x_i}{\Delta x})$, where the coefficients η_i, $i = 0, \ldots, N-1$ are determined such that $\tilde{g}^a(x_i) = g_i$, for $i = 0, \ldots, N-1$.

3.2 The conservative approach

We reconstruct a function $\tilde{g}^c$ which satisfies

$$\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \tilde{g}^c(x)\mathrm{d}x = \bar{g}_i, \quad i = 0, \dots, N-1, \tag{3}$$

and we then take

$$\mathcal{T}_\alpha(\bar{g}_0, \dots, \bar{g}_{N-1})_i = \frac{1}{\Delta x}\int_{x_{i-1/2}+\alpha}^{x_{i+1/2}+\alpha} \tilde{g}^c(x)\mathrm{d}x. \tag{4}$$

We have again several possibilities for the reconstruction of the function $\tilde{g}^c$. We suppose here also periodic boundary conditions, that is $\bar{g}_i = \bar{g}_{i+kN}$ for $k \in \mathbb{Z}$.
Lagrange type interpolation of order $d = 2p$ ($\mathbf{CLag_d}$) : if $x \in (x_{k-1/2}, x_{k+1/2})$, we take $\tilde{g}^c(x) = P(x)$, where P is the unique polynomial of degree $\leq d$ satisfying

$$\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} P(x)\mathrm{d}x = \bar{g}_i, \text{ for } i = k-p, \dots, k+p. \tag{5}$$

Parabolic splines ($\mathbf{CSPL_2}$) : it has already been previously described as the PSM method without limiters. We shortly recall one of its definition, in this context of notations. We define $\tilde{g}^c(x) = P_i(x)$, for $x \in (x_{i-1/2}, x_{i+1/2})$, where the polynomials $P_i, \quad i = 0, \dots, N-1$ are of degree ≤ 2 uniquely determined by the conditions $\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} P_i(x)\mathrm{d}x = \bar{g}_i, \quad i = 0, \dots, N-1$, together with the continuity conditions

$$P_i(x_{i+1/2}) = P_{i+1}(x_{i+1/2}), \quad P_i'(x_{i+1/2}) = P_{i+1}'(x_{i+1/2}), \quad i = 0, \dots, N-1.$$

Note that CLag_2 is in fact the PFC method without limiters [9] and, as we have mentionned it just before, that CSPL_2 is PSM without limiters [26, 27]. In the sequel we will make the link between the conservative and advective approaches and we will prove that CLag_2 is equivalent to LAG3 and that CSPL_2 is equivalent to SPL. We underline that this equivalence only holds for uniform meshes in a constant advection case.

3.3 Equivalence between conservative and advective approach

One interesting property is the following: the schemes deriving from $\mathrm{CLag}_{2\mathrm{p}}$ or from $\mathrm{Lag}_{2\mathrm{p}+1}$ give rise to the same transport operator, and the same is true for CSPL_2 and SPL. We underline that this property only holds in this particular case of a constant $1D$ advection with a uniform mesh. On the other hand, we can make an equivalence between a conservative and an advective approach for a broad range of reconstructions. In the sequel, we will formalize this property and then check on examples for which we have the equivalence $\mathrm{CLag}_{2\mathrm{p}} \sim \mathrm{Lag}_{2\mathrm{p}+1}$ and $\mathrm{CSPL}_2 \sim \mathrm{SPL}$.
We suppose for this that the transport operator has the following form :

$$\mathcal{T}_{\alpha,N}(g_0, \dots, g_{N-1})_i = \sum_{\ell=-s}^{s} L_\ell(x_i + \alpha - x_k)g_{k+\ell}, \tag{6}$$

where k is determined such that $x_i + \alpha$ falls in $[x_k, x_{k+1}[$ and $s \in \mathbb{N}^*$. The number s may depend on N. As examples, we will see that the reconstructions $\mathrm{Lag}_{2\mathrm{p}+1}$ and SPL fall in that category.

From advective to conservative. Now, if the transport operator comes from an advective approach, the reconstruction $\tilde{g}^a$ satisfies for $0 \leq \alpha < \Delta x$

$$\tilde{g}^a(x_i + \alpha) = \sum_{\ell=-s}^{s} L_\ell(\alpha) g_{i+\ell}$$

In particular, we have from the interpolation property (1), by taking $\alpha = 0$,

$$g_i = \sum_{\ell=-s}^{s} L_\ell(0) g_{i+\ell},\ i = 0, \ldots, N-1,$$

and thus we have

$$L_0(0) = 1, \quad L_\ell(0) = 0, \text{ if } \ell \neq 0. \tag{7}$$

In order to make the link with the conservative approach, we want to define a reconstruction $\tilde{g}^c$ which satisfies (3) and which gives rise to the same transport operator. For this, we define for $\alpha \in [0, \Delta x[$ and $i = 0, \ldots, N-1$,

$$G(x_{i+1/2} + \alpha) = \Delta x \sum_{\ell=-s}^{s} L_\ell(\alpha) \left(\sum_{r=-s}^{i+\ell} g_r \right) \tag{8}$$

and

$$\tilde{g}^c(x) = G'(x),\ x_{1/2} \leq x < x_{N+1/2}, \tag{9}$$

and we prolongate $\tilde{g}^c$ on $\mathbb{R}$ by periodicity. We then have for $i = 1, \ldots, N$, by using (7)

$$\frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \tilde{g}^c(x) \mathrm{d}x = \frac{G(x_{i+1/2}) - G(x_{i-1/2})}{\Delta x}$$
$$= \sum_{\ell=-s}^{s} L_\ell(0) \left(\sum_{r=-s}^{i+\ell} g_r \right) - \sum_{\ell=-s}^{s} L_\ell(0) \left(\sum_{r=-s}^{i+\ell-1} g_r \right) = \sum_{r=-s}^{i} g_r - \sum_{r=-s}^{i-1} g_r = g_i,$$

and for $i = 0$, we get

$$\frac{1}{\Delta x} \int_{x_{-1/2}}^{x_{1/2}} \tilde{g}^c(x) \mathrm{d}x = \frac{1}{\Delta x} \int_{x_{N-1/2}}^{x_{N+1/2}} \tilde{g}^c(x) \mathrm{d}x = g_N = g_0,$$

and thus $\tilde{g}^c$ effectively satisfies (3).

On the other hand, we have for $\alpha \in \mathbb{R}$ and $i = 0, \ldots, N-1$,

$$\frac{1}{\Delta x} \int_{x_{i-1/2}+\alpha}^{x_{i+1/2}+\alpha} \tilde{g}^c(x) \mathrm{d}x = \frac{G(x_{k+1/2} + x_{i-k} + \alpha) - G(x_{k-1/2} + x_{i-k} + \alpha)}{\Delta x}$$
$$= \sum_{\ell=-s}^{s} L_\ell(x_{i-k}+\alpha) \left(\sum_{r=-s}^{k+\ell} g_r \right) - \sum_{\ell=-s}^{s} L_\ell(x_{i-k}+\alpha) \left(\sum_{r=-s}^{k-1+\ell} g_r \right) = \sum_{\ell=-s}^{s} L_\ell(x_{i-k}+\alpha) g_{k+\ell},$$

where k is determined such that $x_i+\alpha$ falls in $[x_k,x_{k+1}[$, and thus, we have

$$\mathcal{T}_{\alpha,N}(g_0,\dots,g_{N-1})=\tilde{g}^a(x_i+\alpha)=\frac{1}{\Delta x}\int_{x_{i-1/2}+\alpha}^{x_{i+1/2}+\alpha}\tilde{g}^c(x)\mathrm{d}x,$$

which means that the conservative reconstruction $\tilde{g}^c$ derived from the advective reconstruction defines the same transport operator.

From conservative to advective. We suppose this time that the transport operator comes from a conservative approach. The reconstruction $\tilde{g}^c$ satisfies

$$\frac{1}{\Delta x}\int_{x_{i-1/2}+\alpha}^{x_{i+1/2}+\alpha}\tilde{g}^c(x)\mathrm{d}x=\sum_{\ell=-s}^{s}L_\ell(x_i+\alpha-x_k)g_{k+\ell},\ x_i+\alpha\in[x_k,x_{k+1}[,$$

together with (3). In order to make the link with the advective approach, we simply define $\tilde{g}^a$ by

$$\tilde{g}^a(x_i+\alpha)=\sum_{\ell=-s}^{s}L_\ell(x_i+\alpha-x_k)g_{k+\ell},\ x_i+\alpha\in[x_k,x_{k+1}[, \tag{10}$$

and it remains to prove that we have (7), which comes straightforwardly from (3).

Equivalence for Lagrange reconstructions ($\mathbf{CLag_{2p}\sim Lag_{2p+1}}$).
From the advective approach, we have explicitely (6), with

$$s=p+1,\ L_\ell(\alpha)=\prod_{j=-p-1,\ j\neq\ell}^{p+1}\frac{\alpha-j\Delta x}{(\ell-j)\Delta x}.$$

We then define $\tilde{g}^c$ by (8)-(9). By putting $P(x)=\tilde{g}^c(x)$, for $x\in(x_{k-1/2},x_{k+1/2})$, we have to check that P is of degree $\leq p$ and satisfies (5). From (8)-(9), we get

$$P(x)=\Delta x\sum_{\ell=-p-1}^{p+1}L'_\ell(x-x_{k-1/2})\left(\sum_{r=-p-1}^{k-1+\ell}g_r\right),$$

and thus P is of degree $\leq p$. We also have for $\ell=k-p,\dots,k+p$

$$\frac{1}{\Delta x}\int_{x_{\ell-1/2}}^{x_{\ell+1/2}}P(x)\mathrm{d}x=\Delta x\sum_{j=-p-1}^{p+1}\big(L_j(x_{\ell+1/2}-x_{k-1/2})-L_j(x_{\ell-1/2}-x_{k-1/2})\big)\left(\sum_{r=-p-1}^{k-1+j}g_r\right).$$

On the right hand side, we have $L_j(x_{\ell-1/2}-x_{k-1/2})=L_j((\ell-k)\Delta x)=0$, if $j\neq\ell-k$, since $\ell-k\in\{-p-1,\dots,p+1\}$, and $L_j(x_{\ell-1/2}-x_{k-1/2})=1$, if $j=\ell-k$. We similarly have $L_j(x_{\ell+1/2}-x_{k-1/2})=L_j((\ell+1-k)\Delta x)=0$, if $j\neq\ell+1-k$, since $\ell+1-k\in\{-p-1,\dots,p+1\}$, and $L_j(x_{\ell+1-1/2}-x_{k-1/2})=1$, if $j=\ell+1-k$. We thus get

$$\frac{1}{\Delta x}\int_{x_{\ell-1/2}}^{x_{\ell+1/2}}P(x)\mathrm{d}x=\Delta x\sum_{r=-p-1}^{k-1+\ell+1-k}g_r-\Delta x\sum_{r=-p-1}^{k-1+\ell-k}g_r=g_\ell,$$

which means that (5) is true.

Equivalence for spline reconstructions ($\mathbf{CSPL_2\sim SPL}$).
From the advective approach, we have

$$\tilde{g}^a(x_i+\alpha)=\sum_{\ell=0}^{N-1}\eta_\ell S(\alpha/\Delta x+i-\ell),$$

and the η_ℓ are determined by the conditions

$$\eta_{\ell-1} + 4\eta_\ell + \eta_{\ell+1} = 6g_\ell,\ \alpha_\ell = \alpha_{\ell \mod N},\ \ell \in \mathbb{Z}.$$

Following [15], we can obtain explicitely for $\ell = 0, \ldots, N-1$

$$\eta_\ell = \sum_{r=0}^{N-1} \frac{u_{N-|r-\ell|-1} + u_{|r-\ell|-1}}{2(1-u_N)} 6g_r,$$

by introducing

$$u_N = \frac{(-2+\sqrt{3})^N + (-2-\sqrt{3})^N}{2},\ v_r = \frac{(-2+\sqrt{3})^{r+1} + (-2-\sqrt{3})^{r+1}}{2\sqrt{3}},\ r = 0, \ldots, N-1.$$

From the periodicity of the sequence (g_i), we get for $\ell \in \mathbb{Z}$

$$\eta_\ell = \sum_{s=-N/2}^{N/2-1} \frac{v_{N-|s|-1} + v_{|s|-1}}{2(1-u_N)} 6g_{\ell+s}.$$

This finally gives

$$\tilde{g}^a(x_i + \alpha) = \sum_{j=-N/2}^{N/2-1} \sum_{p=0}^{N-1} \frac{v_{N-|j|-1} + v_{|j|-1}}{2(1-u_N)} 6S(\alpha/\Delta x + i - p) g_{p+j}$$

An example where the advective approach is not conservative. Let $N = 2p \in \mathbb{N}^*$ be an even number. We consider the reconstruction $\tilde{g}^a$ defined by

$$\tilde{g}^a(x_{2i} + \alpha\Delta x) = g_{2i} + (g_{2i+1} - g_{2i})\alpha + \frac{g_{2i+2} - 2g_{2i+1} + g_{2i}}{2}\alpha(\alpha-1),\ 0 \le \alpha < 2.$$

We then have for $0 \le \alpha < 1$

$$\begin{aligned}
& \sum_{i=0}^{N-1} \mathcal{T}_{\alpha,N}(g_0, \ldots, g_{N-1})_i \\
= & \sum_{i=0}^{p-1} (g_{2i} + (g_{2i+1} - g_{2i})\alpha + \frac{g_{2i+2} - 2g_{2i+1} + g_{2i}}{2}\alpha(\alpha-1) \\
+ & \sum_{i=0}^{p-1} (g_{2i} + (g_{2i+1} - g_{2i})(\alpha+1) + \frac{g_{2i+2} - 2g_{2i+1} + g_{2i}}{2}(\alpha+1)\alpha \\
= & (2 - \alpha - (\alpha+1) + \alpha(\alpha-1) + (\alpha+1)\alpha) \sum_{i=0}^{p-1} g_{2i} \\
+ & (\alpha + (\alpha+1) - \alpha(\alpha-1) - (\alpha+1)\alpha) \sum_{i=0}^{p-1} g_{2i+1} \\
= & (1 - 2\alpha + 2\alpha^2) \sum_{i=0}^{p-1} g_{2i} + (1 + 2\alpha - 2\alpha^2) \sum_{i=0}^{p-1} g_{2i+1},
\end{aligned}$$

so that this quantity can depend on α which is not the case for a conservative approach, and thus we can not find $\tilde{g}^c$ satisfying (4) which yields to the same scheme.

3.4 Application to the Vlasov-Poisson model

As an application of the constant advection case, we are concerned with the Vlasov-Poisson model, the unknown of which $f = f(t,x,v)$ is the electronic distribution function. It depends on the spatial variable $x \in [0,L]$ where $L>0$ is the size of the domain, the velocity direction $v \in \mathbb{R}$ and the time $t \geq 0$. The time evolution of this distribution function is given by the following phase space transport equation, the Vlasov equation

$$\frac{\partial f}{\partial t} + v\partial_x f + E(t,x)\partial_v f = 0, \tag{11}$$

with the initial condition

$$f(0,x,v) = f_0(x,v).$$

The electric field $E(t,x)$ is given by the coupling with the distribution function f through the Poisson equation

$$\partial_x E(t,x) = \rho(t,x) - \rho^i, \quad \int_0^L E(t,x)\mathrm{d}x = 0, \tag{12}$$

where the electronic charge density ρ is given by $\rho(t,x) = \int_{\mathbf{R}} f(t,x,v)dv$ and ρ^i denotes the ion density. In this work, we restrict ourselves to a uniform background of ions which leads to $\rho^i = 1$ after a suitable choice of dimensionless parameters.

In view of finite volumes formulation, it will be convenient to re-write the Vlasov equation into a conservative form

$$\frac{\partial f}{\partial t} + \partial_x(vf) + \partial_v(E(t,x)f) = 0. \tag{13}$$

The Vlasov-Poisson model preserves some physical quantities with time which will be analysed and compared for the different numerical methods. First of all, the Vlasov-Poisson equation preserves the L^p norms for $p \geq 0$

$$\frac{d}{dt}\|f(t)\|_{L^p} = 0. \tag{14}$$

The momentum

$$\frac{d}{dt}\left[\int_0^L \int_{\mathbf{R}} vf(t,x,v)dxdv\right] = 0,$$

and the total energy are also constant in time

$$\frac{d}{dt}\mathcal{E}(t) = \frac{d}{dt}\mathcal{E}_k(t) + \frac{d}{dt}\mathcal{E}_e(t) = \frac{d}{dt}\int_0^L \int_{\mathbf{R}} f(t,x,v)\frac{|v|^2}{2}dxdv + \frac{1}{2}\frac{d}{dt}\int_0^L \int_{\mathbf{R}} |E(t,x)|^2 dxdy, \tag{15}$$

where $\mathcal{E}_e$ and $\mathcal{E}_k$ denote the electric and kinetic energy respectively. From a numerical point of view, the good conservation of these different quantities is an important feature for Vlasov simulations.

3.4.1 The general algorithm

In this subsection, we review the main steps of a semi-Lagrangian method in the case of directional splitting with constant advection, which is applied for the discretization of the Vlasov-Poisson model.

Discretization of the distribution function. The unknown quantities are then $f^n_{k,\ell}$ which are approximations of $f(t^n, x_k, v_\ell)$. We suppose periodic boundary conditions so that we only have to compute at each time t^n

$$f^n_{k,\ell}, \text{ for } k = 0, \dots, N_x - 1, \ \ell = 0, \dots, N_v - 1.$$

Transport operator. Let us define for $\alpha \in \mathbb{R}$ and $N \in \mathbb{N}^*$ a transport operator $\mathcal{T}_{\alpha,N} : \mathbb{R}^N \to \mathbb{R}^N$. We will define several possibilities for the transport operator. If $(f_0, \dots, f_{N-1})$ is a discretization of a function f, then $\mathcal{T}_{\alpha,N}(f_0, \dots, f_{N-1})$ should be a discretization of the shifted function $x \to f(x+\alpha)$. We also denote by $\mathcal{T}^x_\alpha : \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$ (resp. $\mathcal{T}^v_\alpha : \mathbb{R}^{N_v} \to \mathbb{R}^{N_v}$) a transport operator which shifts along the x (resp. v) direction.

Algorithm. The time splitting algorithm then reads

Step 0. Initialization : $f_{k,\ell} = f_0(x_k, v_\ell), \ k = 0, \dots, N_x - 1, \ell = 0, \dots, N_v - 1$.
Step 1. Half time step shift along the x-axis:
For each $\ell = 0, \dots, N_v - 1$, $(f_{k,\ell})_{k=0}^{N_x-1} \to \mathcal{T}^x_\alpha((f_{k,\ell})_{k=0}^{N_x-1})$ with $\alpha = -v_\ell \Delta t/2$.
Step 2. Computation of the charge density and the electric field by integrating (12).
Step 3. Shift along the v-axis:
For each $k = 0, \dots, N_x - 1$, $(f_{k,\ell})_{\ell=0}^{N_v-1} \to \mathcal{T}^v_\alpha((f_{k,\ell})_{\ell=0}^{N_v-1})$ with $\alpha = -E_k \Delta t$.
Step 4.a Half time step shift along the x-axis:
For each $\ell = 0, \dots, N_v - 1$, $(f_{k,\ell})_{k=0}^{N_x-1} \to \mathcal{T}^x_\alpha((f_{k,\ell})_{k=0}^{N_x-1})$ with $\alpha = -v_\ell \Delta t/2$.
Step 4.b We have $f^n_{k,\ell} = f_{k,\ell}$, for $k = 0, \dots, N_x - 1, \ \ell = 0, \dots, N_v - 1$.
Step 4.c Half time step shift along the x-axis:
For each $\ell = 0, \dots, N_v - 1$, $(f_{k,\ell})_{k=0}^{N_x-1} \to \mathcal{T}^x_\alpha((f_{k,\ell})_{k=0}^{N_x-1})$ with $\alpha = -v_\ell \Delta t/2$.
Step 5. $n \to n+1$ and loop to **Step 2.**

Note that if we make no diagnostic of the distribution function, we can simplify **Step 4.a-c** into

Step 4. Shift along the x-axis:
For each $\ell = 0, \dots, N_v - 1$, $(f_{k,\ell})_{k=0}^{N_x-1} \to \mathcal{T}^x_\alpha((f_{k,\ell})_{k=0}^{N_x-1})$ with $\alpha = -v_\ell \Delta t$.

In the sequel, we present numerical results for the Vlasov-Poisson equation for which several choices of transport operators $\mathcal{T}_{\alpha,N}$ are performed.

3.4.2 Numerical results: Bump-on-tail test case

The present numerical schemes are validated on academic test cases introduced in [9, 10, 19]. In the present work, numerical results obtained by the methods of section 2 are applied on the bump-on-tail instability test case for which the initial condition writes (see [20, 19])

$$f_0(x,v) = \tilde{f}(v)[1 + 0.04\cos(kx)], \quad x \in [0, L_x], v \in [-v_{\max}, v_{\max}],$$

with $k = 0.5$, $v_{\max} = 9$, $L_x = 20\pi$; moreover, we have

$$\tilde{f}(v) = n_p \exp(-v^2/2) + n_b \exp\left(-\frac{|v-u|^2}{2v_t^2}\right)$$

whose parameters are

$$n_p = \frac{9}{10(2\pi)^{1/2}}, n_b = \frac{2}{10(2\pi)^{1/2}}, u = 4.5, v_t = 0.5.$$

The numerical parameters are $N_x = 128, N_v = 128, \Delta t = 0.1$.

For this test case, we are interested in the time evolution of the total energy $\mathcal{E}$ together with the electric energy $\mathcal{E}_e$ given by (15). We also look after the $L^p, p = 1, 2$ norms of f (see (14)) which are conserved with time. The same is true for the total energy whereas the electric energy has an oscillatoring behaviour for large times (see [19]). Three vortices are then created in the phase space which are moving along the velocity $v = v_t$ (BGK equilibrium) and a lack of accuracy leads to a vorticies merging which has for consequence a loss of the oscillatoring behaviour of the electric energy.

Our goal is to compare the different methods we talked about using a Strang splitting (see the previous algorithm): the conservative methods PFC of [9], PFC2 (using the filter (19), PSM (using the filter (21)), PSM ZER of [26] and the advective methods LAG3 and SPL. Note that the Forward Update method introduced in section 2.2 is not shown in the present case; indeed the displacements are constant so that only uniform mesh are generated by the characteristics and the method is completely equivalent to the SPL or the PSM methods.

Fig. 3 shows the time evolution of the electric energy for the different methods. The main features (see [20, 19]) of the expected behaviour is respected by all the methods: the electric energy presents a maximum at $t \approx 20\,\omega_p^{-1}$ and then an slowing oscillatoring behaviour on which is superimposed the oscillation of the system at ω_p. Nevertheless, two classes can be distinguished: splines and Lagrange based methods. For the methods based on Lagrange interpolation (LAG3, PFC), the oscillations of the electric energy due to the particles trapping are damped and the amplitude is decreasing for large time. It is not the case for the splines based methods (SPL, PSM) which keep the slow oscillatoring behaviour and a constant amplitude up to the end of the simulation. This can be explained by the fact that fine structures are developed in the vortices; they are quickly eliminated by the methods based on Lagrange interpolation whereas splines methods follow these thin details of the phase space solution for longer times. Hence, methods based on Lagrange interpolation seem to be not sufficiently accurate to describe this kind of phenomenon.

In Fig. 4, the $L^p, p = 1, 2$ norms are plotted with respect to time for LAG3, PFC, PFC2, SPL and PSM. As expected, the conservative methods (PFC, PFC2, PSM) which can be coupled with a positive filter ensure the positivity of the solution so that the L^1 norm is preserved. It is not the case for the advective semi-Lagrangian methods LAG3 and SPL. In particular, LAG3 gives rise to bad results and the filters of the PFC methods has to act often to ensure positivity of the reconstruction. The SPL method also generates negative values, but in a reasonable proportion so that PSM filters do not intervene often. The two classes of methods are emphasized by the time evolution of the L^2 norm. Methods based on Lagrange interpolation (which are nearly superimposed) present a more diffusive behaviour compared to spline based methods. The smoothing effects of the methods based on Lagrange interpolation lead to a bad conservation of the L^2 norm since they do not follow the filamentation of the distribution function for long times.

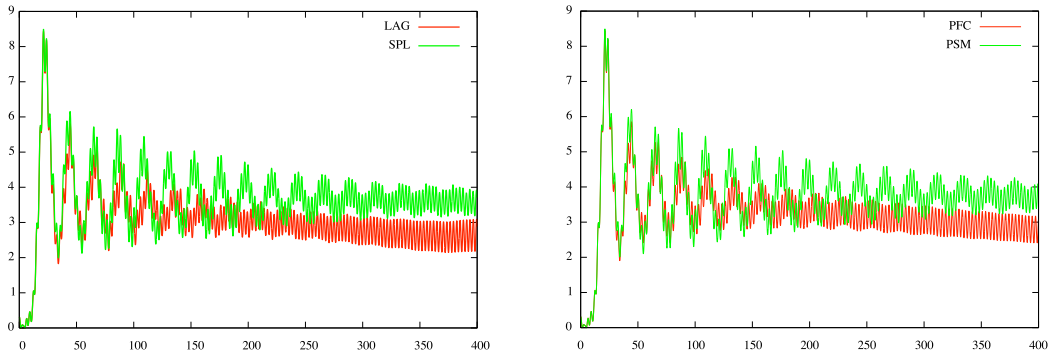

Figure 3: Time evolution of the electric energy for the different methods: semi-Lagrangian methods LAG3 and SPL (left) and conservative methods PFC and PSM (right). $N_x = N_v = 128, \Delta t = 0.1$ for the Bump-on-tail test.

In Fig. 5, we distinguish the splines based methods (left figure) from the Lagrange ones (right figures). We add on the left figure the numerical results obtain by the PSM ZER method coupled with a monotone filter introduced in [26]. This filter is not very appropriate in our context and the optimal filter reconstructs a solution which is closer to the basic cubic splines solution.

Finally, in Fig. 6, the time evolution of the total energy is plotted for the different methods. This quantity is quite difficult to preserve at the discrete level (see [24, 4]). In particular, it is very difficult to ensure both positivity of the solution and conservation of the total energy for nonlinear tests. Numerical results are presented in Fig. 6; we first notice that no method preserves exactly this quantity. Moreover, we can observe the influence of the slope limiters on the results: indeed, the difference between PFC and Lagrange method is roughly the same as for the L^1 time evolution (see Fig. 4): PFC ensures positivity to LAG3 but it is to the detriment of the total energy conservation. This figure also emphasizes the good behaviour of the optimal filter applied to PSM. The conservation of the total energy is lower than 0.1%, which is similar to the SPL method.

As a conclusion, the PSM filter improves the numerical results of the original PSM method for every conserved quantity. The loss of total energy conservation is about a 0.1%, which remains very reasonable. Let us remark that a time discretization refinement has a great influence on the conservation of the total energy.

4 Non-constant case: the guiding-center model

In this work, we also deal with another type of Vlasov equation for which the advection term is not constant. The so-called guiding-center model enters in this category (see [22]). This model, which has been derived to describe highly magnetized plasma in the transverse plane of a tokamak, considers the evolution of the particles density $\rho(t, x, y)$

$$\partial_t \rho + E^\perp \cdot \nabla \rho = 0, \tag{1}$$

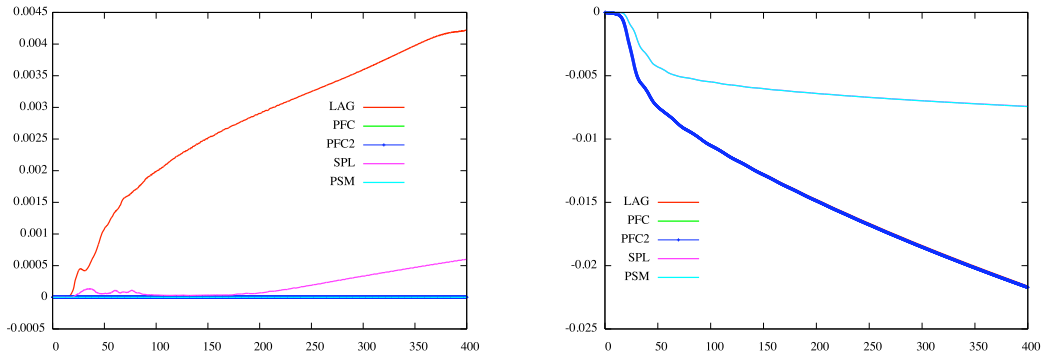

Figure 4: Time evolution of the L^1 norm (left) and L^2 norm (right) for LAG3, PFC, PFC2, SPL and PSM. $N_x = N_v = 128, \Delta t = 0.1$ for the Bump-on-tail test.

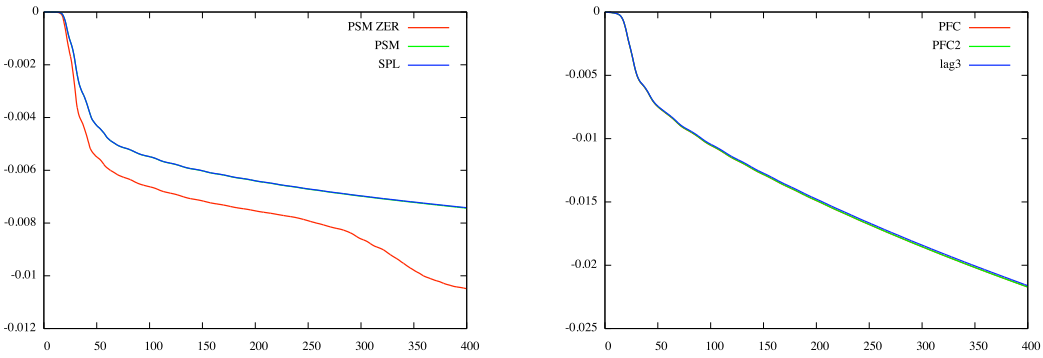

Figure 5: Time evolution of the L^2 norm for spline based methods PSM ZER (with monotone filter of [26]), PSM (optimal filter) and SPL (left) and methods based on Lagrange interpolation PFC, PFC2 (PFC with optimal filter) and LAG3 (right). $N_x = N_v = 128, \Delta t = 0.1$ for the Bump-on-tail test.

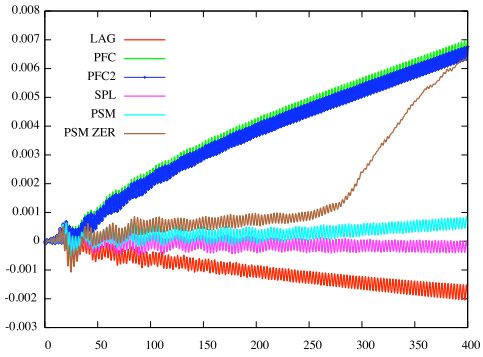

Figure 6: Time evolution of the total energy for the different methods. $N_x = N_v = 128, \Delta t = 0.1$ for the Bump-on-tail test.

where the electric field

$$E = E(x,y) = (E_x(x,y), E_y(x,y)),$$

satisfies a Poisson equation

$$-\Delta\Phi = \rho, \quad E = -\nabla\Phi. \tag{2}$$

We denote by $E^\perp = (E_y, -E_x)$. The specificity of (1) lies on the fact that one-dimensional splitting cannot (in principle) be applied (see [22, 14]) since the advection term $E^\perp$ depends on (x,y). Consequently, this model contains additional difficulties compared to the Vlasov-Poisson model and seems to be a good candidate to test numerical methods.

To that purpose, we briefly recall the conservation properties of (1) which should be preserved in the best manner by the numerical schemes. The guiding center model (1) preserves the total mass, the L^2 norm of the density (enstrophy) and the L^2 norm of the electric field (energy)

$$\frac{d}{dt}\iint \rho(t,x,y)dxdy = \frac{d}{dt}\|\rho(t)\|_{L^2} = \frac{d}{dt}\|E(t)\|_{L^2} = 0. \tag{3}$$

4.1 The general algorithm

In this subsection, we review the main steps of a semi-Lagrangian method in the case of directional splitting which is applied for the discretization of the guiding-center-Poisson model.

Grid notations. Let $N_x, N_y \in \mathbb{N}^*$, $y_{\max} > 0$, a time step $\Delta t > 0$.
We define then classically as notations

$$\Delta x = L_x/N_x,\ \Delta y = L_y/N_y\ x_k = kL_x/N_x,\ y_\ell = \ell L_y/N_y$$

for $k = 0, \ldots, N_x,\ \ell = 0, \ldots, N_y$ and $t^n = n\Delta t,\ n \in \mathbb{N}$.

Discretization of the distribution function. The unknown quantities are then $\rho^n_{k,\ell}$ which are approximations of $\rho(t^n, x_k, y_\ell)$. We suppose periodic boundary conditions so that we only have to compute at each time t^n

$$\rho^n_{k,\ell},\ \text{for } k = 0, \ldots, N_x - 1,\ \ell = 0, \ldots, N_y - 1.$$

Transport operator. Let us define for $(\alpha_k) \in \mathbb{R}^{N_x+1}$ a transport operator $\mathcal{T}_\alpha : \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$. For the conservative approaches we detailed in section 2.1, this operator writes

$$\mathcal{T}_\alpha(\bar\rho_0, \bar\rho_1, ..., \bar\rho_{N_x-1}) = \left(\frac{1}{\Delta x}\int_{x_{k-1/2}-\alpha_{k-1/2}}^{x_{k+1/2}-\alpha_{k+1/2}} \bar\rho(x)dx\right)_{k=0,..,N-1}.$$

The sequence α is determined following one of the algorithms detailed in sections 2.1.1 and 2.2.1.

Algorithm

Step 0. Initialization : $\rho_{k,\ell} = \rho_0(x_k, y_\ell),\ k = 0, \ldots, N_x - 1, \ell = 0, \ldots, N_y - 1$.
Step 1. Compute of the electric field (E_x^0, E_y^0) by integrating (2).
Step 2. Compute $\rho_{k,\ell}^1$ using ρ^0:
Step 2.a. Half time step shift along the x-axis:
Compute the x-displacement for each ℓ $\alpha_k = \Delta t/4 E_y^0(x_k - \alpha_k, y_\ell)$
For each $\ell = 0, \ldots, N_y - 1$, $(\rho_{k,\ell})_{k=0}^{N_x-1} \to \mathcal{T}_\alpha^x((\rho_{k,\ell})_{k=0}^{N_x-1})$.
Step 2.b. Shift along the y-axis:
Compute the y-displacement for each k $\alpha_\ell = -\Delta t/2 E_x^0(x_k, y_\ell - \alpha_\ell)$
For each $k = 0, \ldots, N_x - 1$, $(\rho_{k,\ell})_{\ell=0}^{N_y-1} \to \mathcal{T}_\alpha^y((\rho_{k,\ell})_{\ell=0}^{N_y-1})$.
Step 2.c. Half time step shift along the x-axis:
Compute the x-displacement for each k $\alpha_k = \Delta t/4 E_y^0(x_k - \alpha_k, y_\ell)$
For each $\ell = 0, \ldots, N_y - 1$, $(\rho_{k,\ell})_{k=0}^{N_x-1} \to \mathcal{T}_\alpha^x((\rho_{k,\ell})_{k=0}^{N_x-1})$.
Step 3. Compute the electric field (E_x^1, E_y^1) by integrating (2).
Step 4. Compute $\rho_{k,\ell}^{n+1}$ using ρ^{n-1}, ρ^n:
Step 4.a. Half time step shift along the x-axis:
Compute the x-displacement for each ℓ $\alpha_k = \Delta t/2 E_y^n(x_k - \alpha_k, y_\ell)$
For each $\ell = 0, \ldots, N_y - 1$, $(\rho_{k,\ell})_{k=0}^{N_x-1} \to \mathcal{T}_\alpha^x((\rho_{k,\ell})_{k=0}^{N_x-1})$.
Step 4.b. Shift along the y-axis:
Compute the y-displacement for each k $\alpha_\ell = -\Delta t E_x^n(x_k, y_\ell - \alpha_\ell)$
For each $k = 0, \ldots, N_x - 1$, $(\rho_{k,\ell})_{\ell=0}^{N_y-1} \to \mathcal{T}_\alpha^y((\rho_{k,\ell})_{\ell=0}^{N_y-1})$.
Step 4.c. Half time step shift along the x-axis:
Compute the x-displacement for each ℓ $\alpha_k = \Delta t/2 E_y^n(x_k - \alpha_k, y_\ell)$
For each $\ell = 0, \ldots, N_y - 1$, $(\rho_{k,\ell})_{k=0}^{N_x-1} \to \mathcal{T}_\alpha^x((\rho_{k,\ell})_{k=0}^{N_x-1})$.
Step 5. Compute the the electric field (E_x^{n+1}, E_y^{n+1}) by integrating (2).
Step 6. $n \to n + 1$ and loop to **Step 4.**

Different methods will be compared: PSM without filter, the forward Update Method (FUM), (detailed in subsection 2.2) and the traditional semi-Lagrangian method with cubic splines interpolation developed in [22]. For this last approach, we look after the validity of the splitting. Indeed, as we already discussed in section 2, this method is not always conservative when the splitting procedure is performed (the method will be refered as SPL1D). However, if a full two-dimensional interpolation is performed, the method becomes conservative (the method will be refered as SPL2D). Hence, this point will be discussed.

4.2 Numerical results: Kelvin-Helmoltz instability test case

We consider the Kelvin-Helmoltz instability in the periodic-periodic case for which the growth rate of the instability can be computed *a priori*. This is of great importance to check at the qualitative point of view the accuracy of the code.

Following the computations of [21], the linearization (1)-(2) leads to the so-called stability Rayleigh equation. Considering as initial condition a periodic perturbation of the equilibrium solution to (1)-(2), it is possible to start a Kelvin-Helmoltz instability. The difference between the Dirichlet-periodic case (which has been solved in [22]) occurs in the neutrally stable solution which is

equal to 1 in our case (instead of $\sin(y/2)$ in the Dirichlet-periodic case). Then, we can deduce the initial condition for (1)-(2),

$$\rho(x, y, t = 0) = \sin(y) + \varepsilon \cos(kx),$$

where $k = 2\pi/L_x$ is the wave number associated to the length L_x of the domain in the x-direction. The size of the domain in the y-direction is $L_y = 2\pi$. Shoucri's analysis predicts an instability when k is chosen lower to 1. Otherwise, the initial perturbation remains unchanged, neither damped (since (1)-(2) is only a fluid model, not a kinetic model), neither increased.

Various approaches can be employed to determine the instability growth rate of the chosen mode k. A finite difference numerical scheme has been applied to approximate the stability Rayleigh equation which leads to a eigenvalue problem. The results obtained by this way are very closed to those obtained by numerically solve the linearized problem as performed in [21].

The numerical parameters are chosen as follows:

$$k = 0.5, N_x = N_y = 128, \text{ and } \Delta t = 0.1.$$

Let us recall that periodic conditions are considered here; even if the present test bears similarities with the Dirichlet-periodic test presented in [22], the dynamics of the unknown is quite different in the present periodic-periodic context.

For this test case, we are interested in the time evolution of the conserved quantities (3). We also look carefully at the conservation of the total mass in order to verify the difference between conservative and non-conservative methods. We focus our comparisons on splines based methods which appear to be more competitive for strongly nonlinear problems: SPL (SPL1D refers to non-conservative splitting and SPL2D refers to the full two-dimensional advection without splitting), FUM (Forward Update Method presented in subsection 2.2), and PSM (with splitting). As a diagnostics, it is also interesting to look after the $2D$ unknown to realize the fine structures developed along the simulation.

In Fig. 7, the time history of the total mass and the L^2 norm is plotted for the different methods. First, as discussed in [14, 18, 23], SPL1D does not preserve exactly the total mass whereas other methods do. This is expected since this approach solves the non-conservative form of the equation which is not appropriate with the splitting procedure. Then, we observe that the conservative methods present very similar behaviour compared to the method of reference SPL2D: they are conservative and the decay of the L^2 norm occurs at $t \approx 30\,\omega_p^{-1}$. This decay corresponds to the saturation of the instability. Very fine structures are created which can not be captured by the numerical schemes since their size becomes smaller than the grid size.

In Fig. 8, the logarithm of the first Fourier mode of the electric field E_x is plotted as a function of time. The linear theory predicts an exponential growing, the rate of which can be computed *a priori* by solving an eigenvalue problem. This can be performed and the results can be compared to the numerical results. The numerical growth rate corresponds to the slope of the straight line which approximates the logarithm of the first Fourier mode of E_x in the linear phase (between $t \approx 5\,\omega_p^{-1}$ and $t \approx 10\,\omega_p^{-1}$). Considering different values of the wave number k, it is possible to plot the quantity ω/k (where ω is the growth rate of the first Fourier mode of E_x) as a function of $(k_s - k)$ where $k_s = 1$ in our case ($k_s = \sqrt{3}/2$ in the Dirichlet-periodic case). This is performed in Fig. 9

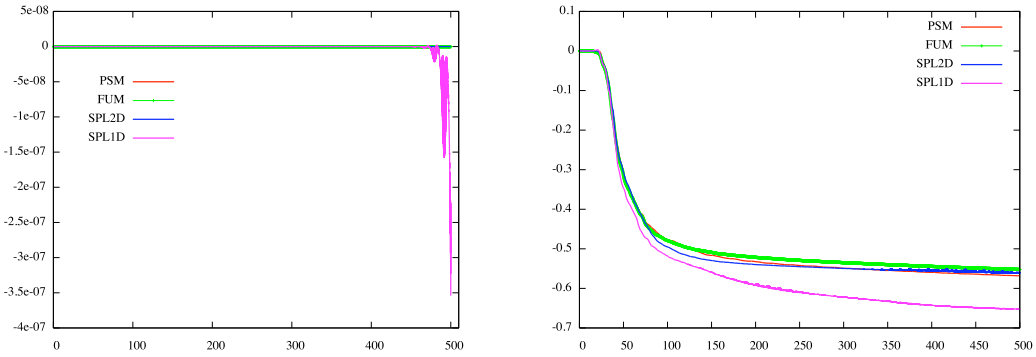

Figure 7: Time evolution of the total mass and the enstrophy for SPL (with splitting (SPL1D) and without splitting (SPL2D)), FUM and PSM. $N_x = N_y = 128, \Delta t = 0.1$ for the Kelvin-Helmoltz instability test.

(right); we can observe the very good agreement between the analytical and the numerical values. This kind of validation is of great importance since a quantitative comparison can be performed, at least on the linear phase.

On Fig. 9 (left), the L^2 norm of the electric field is plotted as a function of time. This quantity is preserved with time by the continuous model. The conservative and splitting procedure based methods present a very good conservation of the energy whereas it is not the case of the non-conservative method SPL1D. The method SPL2D does not preserve very well the energy compared to FUM or PSM for example.

Finally, on Fig. 10 and Fig. 11 we plot the distribution function for the different methods at time $t = 30$ and $t = 60\omega_p^{-1}$. These results confirm the previous observations: first, the PFC method is diffusive (the thin structures are smoothed) and the SPL1D scheme leads to a bad behaviour since the main structures are not respected. In contrast, the FUM and PSM present a good behaviour, very similar to SPL2D.

As a conclusion, the conservative methods present a very good behaviour on this strongly nonlinear and large time case. The splitting procedure also enables to save memory since one-dimensional structures are often used (instead of two-dimensional structures for the computation of the cubic spline coefficients in SPL2D for example). On the other side, we remarked that SPL2D seems to be able to support larger time steps, but we think that the implementation of high order numerical scheme in time could stabilize PSM or FUM when large time steps are used. This extension will be studied in a future work.

5 Conclusion

In this work, new conservative methods have been introduced and then compared to existing methods for equations occuring in plasma physics. Several properties make them very competitive. On the one side, their inherent conservation property enables the use of splitting procedure, which makes easier the implementation of multi-dimensional problems. On the other side, slope limiters can be introduced to ensure the positivity of the unknown.

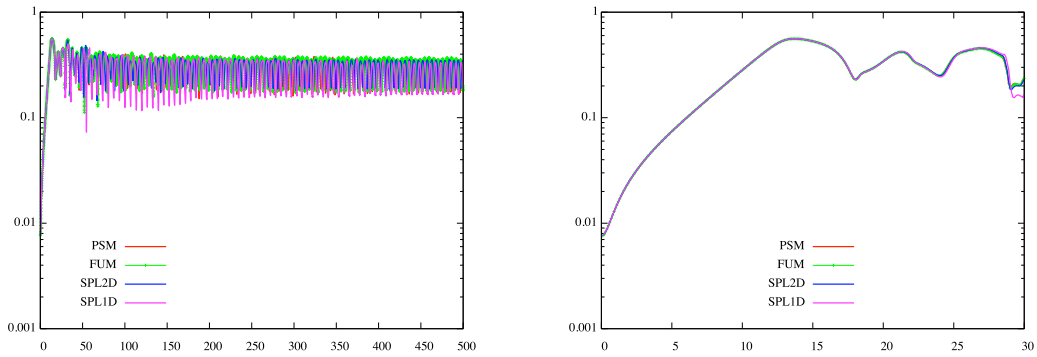

Figure 8: Time evolution of the logarithm of the first Fourier mode for SPL (with splitting (SPL1D) and without splitting (SPL2D)), FUM and PSM. A zoom has been applied on the right figure. $N_x = N_y = 128, \Delta t = 0.1$ for the Kelvin-Helmoltz instability test.

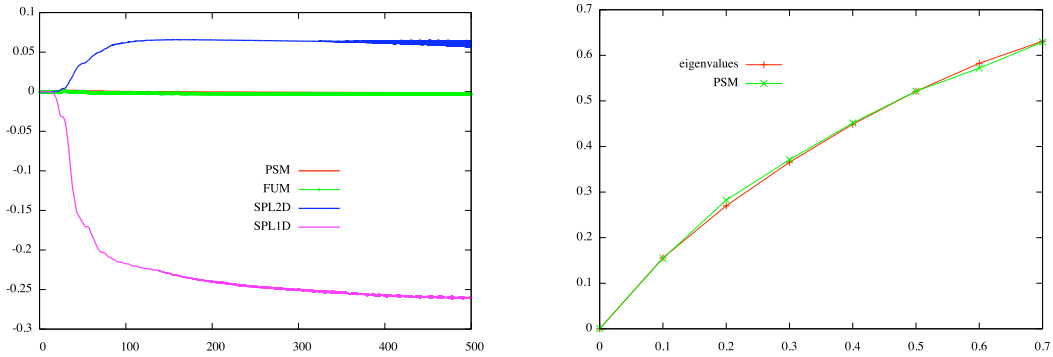

Figure 9: Left figure: Time evolution of the energy for SPL (with splitting (SPL1D) and without splitting (SPL2D)), FUM and PSM. Right figure: normalized growth rate ω/k as a function of $1 - k$. $N_x = N_y = 128, \Delta t = 0.1$ for the Kelvin-Helmoltz instability test.

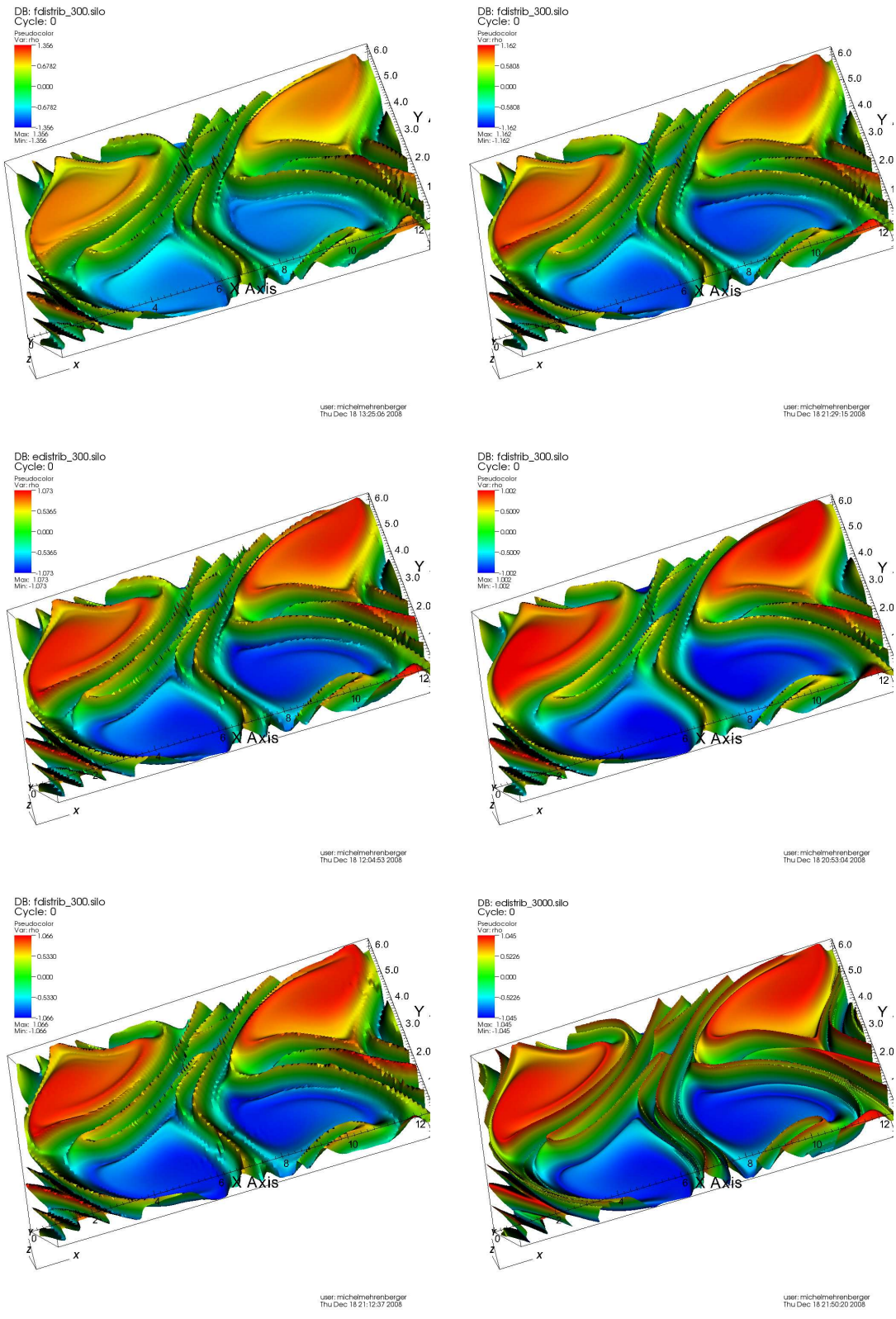

Figure 10: Distribution function for the Kelvin-Helmoltz instability test at time $t = 30\omega_p^{-1}$. Respectively for top-left to bottom-right: PSM, FUM, SPL2D, PFC, SPL1D with $N_x = N_y = 128, \Delta t = 0.1$ and SPL2D with $N_x = N_y = 512, \Delta t = 0.01$.

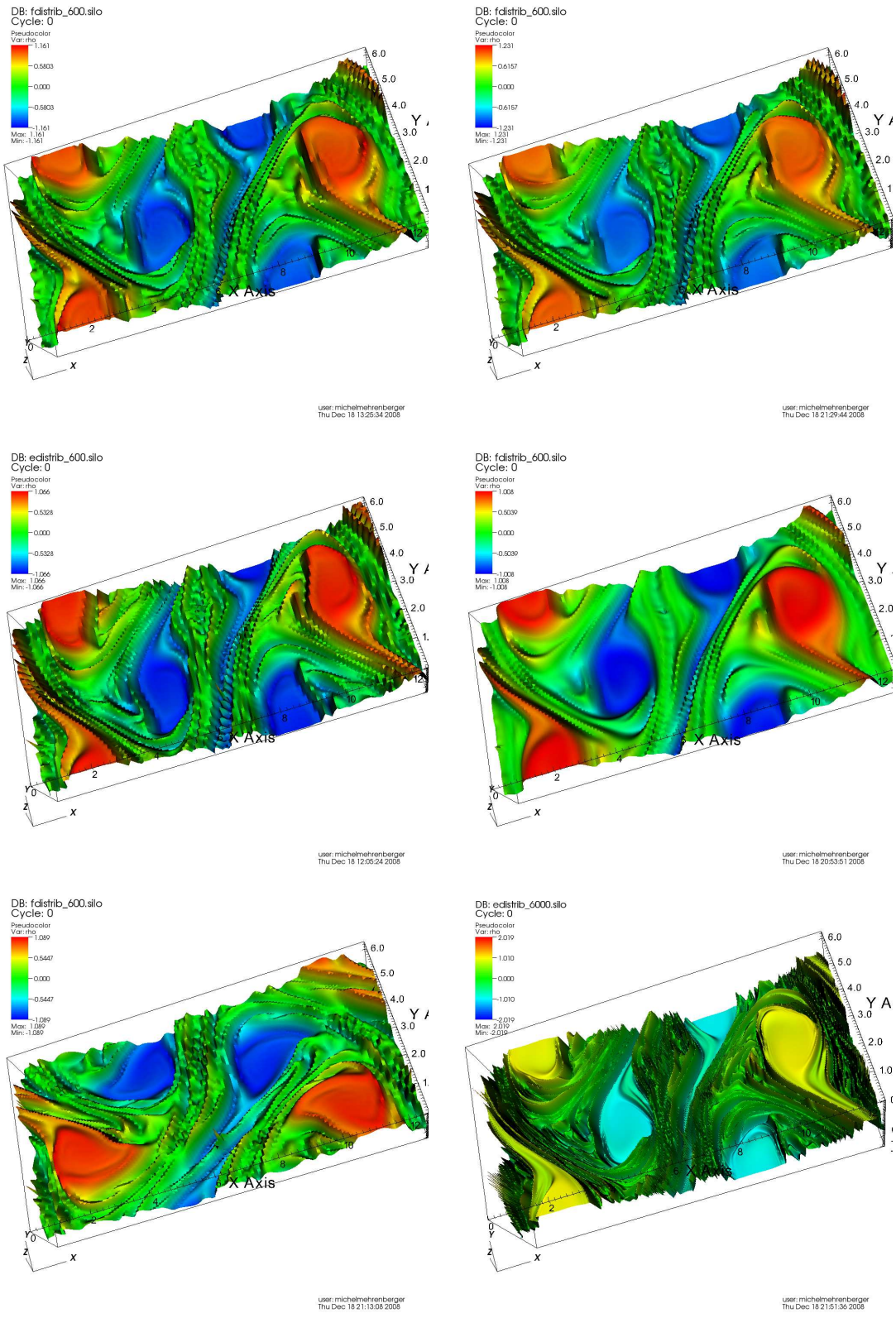

Figure 11. Distribution function for the Kelvin-Helmoltz instability test at time $t = 60\omega_p^{-1}$. Respectively for top-left to bottom-right: PSM, FUM, SPL2D, PFC, SPL1D with $N_x = N_y = 128, \Delta t = 0.1$ and SPL2D with $N_x = N_y = 512, \Delta t = 0.01$.

When they are compared to existing semi-Lagrangian methods, we first observed that for the guiding-center problem, as expected, the advective approaches lead to inaccurate results when splitting procedure is applied. This is not the case for conservative methods. Moreover, they are at least as accurate as the reference methods (SPL2D). For simpler cases like the Vlasov-Poisson model in which the advection term is constant, we proved that the advective methods and their conservative counterparts are equivalent. Obviously, this does not remain true in the non-constant advection case (like for the guiding-center model) which often occurs in particular in gyrokinetic models. The extension of the PSM method to such multi-dimensional system is currently investigated.

Moreover, the splitting procedure makes easier the use of high order numerical scheme in time for backward and forward approaches. The use of high order time splittings (see [25]) is also under investigations.

References

[1] N. Besse, M. Mehrenberger, *Convergence of classes of high order semi-Lagrangian schemes for the Vlasov-Poisson system*, Math. Comput., **77**, pp. 93-123, (2008).

[2] C.K. Birdsall, A. B. Langdon, *Plasma Physics via Computer Simulation*, Inst. of Phys. Publishing, Bristol/Philadelphia, 1991.

[3] M. Brunetti, X. Lapillonne, S. Brunner, T. M. Tran, *Comparison of semi-Lagrangian and Eulerian algorithms for solving Vlasov-type equations*, colloque numérique suisse, 12 avril 2008, EPFL.

[4] J.-A. Carillo, F. Vecil, *Non oscillatory interpolation methods applied to Vlasov-based models*, SIAM J. of Sc. Comput., **29**, pp. 1179-1206, (2007).

[5] P. Colella, P. R. Woodward, *The piecewise parabolic method (PPM) for gas-dynamical simulations*, J. Comput. Phys., **54**, 174-201, (1984).

[6] C. Z. Cheng, G. Knorr, *The integration of the Vlasov equation in configuration space*, J. Comput. Phys, **22**, pp. 330-3351, (1976).

[7] N. Crouseilles, T. Respaud, E. Sonnendrücker, *A forward semi-Lagrangian scheme for the numerical solution of the Vlasov equation*, INRIA research report 6727 (2008).

[8] C. de Boor, *A practical guide to splines*, Springer-Verlag, 1978.

[9] F. Filbet, E. Sonnendrücker, P. Bertrand, *Conservative numerical schemes for the Vlasov equation*, J. Comput. Phys., **172**, pp. 166-187, (2001).

[10] F. Filbet, E. Sonnendrücker, *Comparison of Eulerian Vlasov solvers*, Comput. Phys. Comm., **151**, pp. 247-266, (2003).

[11] A. Ghizzo, P. Bertrand, M.L. Begue, T.W. Johnston, M. Shoucri, *A Hilbert-Vlasov code for the study of high-frequency plasma beatwave accelerator*, IEEE Transaction on Plasma Science, **24**, p. 370, (1996).

[12] V. Grandgirard, M. Brunetti, P. Bertrand, N. Besse, X. Garbet, P. Ghendrih, G. Manfredi, Y. Sarrazin, O. Sauter, E. Sonnendrücker, J. Vaclavik, L. Villard, *A drift-kinetic semi-Lagrangian 4D code for ion turbulence simulation*, J. Comput. Phys., **217**, pp. 395-423, (2006).

[13] F. Coquel, Ph. Helluy, J. Schneider, *Second order entropy diminishing scheme for the Euler equations*, Intern. J. Numer. Meth. in Fluids, **50**, pp. 1029-1061, (2006).

[14] F. Huot, A. Ghizzo, P. Bertrand, E. Sonnendrücker, O. Coulaud, *Instability of the time splitting scheme for the one-dimensional and relativistic Vlasov-Maxwell system*, J. Comput. Phys., **185**, pp. 512-531, (2003).

[15] D. Kershaw, *The explicit inverse of two commonly occuring matrices*, Math. of Comp. **23**, pp. 189-191, (1969).

[16] J. Laprise, A. Plante *A class of semi-Lagrangian integrated-mass (SLIM) numerical transport algorithms*, Mon. Wea. Rev. **123**, pp. 553-565, (1995).

[17] P. H. Lauritzen *An inherently mass-conservative semi-implicit semi-Lagrangian model*, PhD thesis, Department of Geophysics, University of Copenhagen, Denmark, September, 2005.

[18] T. Nakamura, R. Tanaka, T. Yabe, K. Takizawa, *Exactly conservative semi-Lagrangian scheme for multi-dimensional hyperbolic equations with directional splitting technique*, J. Comput. Phys., **174**, pp. 171-207, (2001).

[19] T.Nakamura, T.Yabe, *Cubic Interpolated Propagation Scheme for Solving the Hyper-Dimensional Vlasov-Poisson Equation in Phase Space*, Comput. Phys. Comm., **120**, pp.122-154 (1999).

[20] M. Shoucri, *Nonlinear evolution of the bump-on-tail instability*, Phys. Fluids, **22**, p. 2038, (1979).

[21] M. Shoucri, *A two-level implicit scheme for the numerical solution of the linearized vorticity equation*, Int. J. Numer. Meth. Eng. **17**, p. 1525, (1981).

[22] E. Sonnendrücker, J. Roche, P. Bertrand, A. Ghizzo *The semi-Lagrangian method for the numerical resolution of the Vlasov equation,* J. Comput. Phys., **149**, pp. 201-220, (1999).

[23] R. Tanaka, T. Nakamura, T. Yabe, *Constructing exactly conservative scheme in a non-conservative form*, Comput. Phys. Comm., **126**, pp. 232-243, (2000).

[24] T. Umeda, M. Ashour-Abdalla, D. Schriver, *Comparison of numerical interpolation schemes for one-dimensional electrostatic Vlasov code*, J. Plasma Phys., **72**, pp. 1057-1060, (2006).

[25] H. Yoshida, *Construction of higher order symplectic integrators*, Phys. Lett. A, **150**, p. 262, (1990).

[26] M. Zerroukat, N. Wood, A. Staniforth, *A monotonic and positive-definite filter for a Semi-Lagrangian Inherently Conserving and Efficient (SLICE) scheme*, Q.J.R. Meteorol. Soc., **131**, pp 2923-2936, (2005).

[27] M. Zerroukat, N. Wood, A. Staniforth, *The Parabolic Spline Method (PSM) for conservative transport problems*, Int. J. Numer. Meth. Fluids, **51**, pp. 1297-1318, (2006).

Contents

1 Introduction **3**

2 Conservative methods **4**
2.1 Conservative semi-Lagrangian methods for one-dimensional problems 5
2.1.1 Computation of the characteristic curves 6
2.1.2 Formulation by primitive 7
2.2 Forward update of the characteristics 9
2.2.1 Computation of the characteristics curves 9
2.2.2 Interpolation using a non-uniform mesh 10
2.3 Slope limiters 10

3 The constant advection case **13**
3.1 The advective approach 13
3.2 The conservative approach 14
3.3 Equivalence between conservative and advective approach 14
3.4 Application to the Vlasov-Poisson model 18
3.4.1 The general algorithm 19
3.4.2 Numerical results: Bump-on-tail test case 19

4 Non-constant case: the guiding-center model **21**
4.1 The general algorithm 23
4.2 Numerical results: Kelvin-Helmoltz instability test case 24

5 Conclusion **26**

Centre de recherche INRIA Nancy – Grand Est
LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399