Observable Error in Conservative Dynamical Low-rank Integration Schemes for the Vlasov-Poisson Equation

April Herwig

March 19, 2025

Abstract

Vlasov-Poisson has (infinitely) many conserved quantities. Low-rank tensor approximation [5]. Does this conserve more observables than we expect?

Contents

1 Introduction

Efficient numerical solutions to kinetic equations is the basis for many fields in physics and has applications such as fusion energy generation in magnetically confined chambers, or dosage calculation in radiation therapy. Kinetic models are dependent on both physical space coordinates $x \in \Omega_x \subset \mathbb{R}^d$, $d = 1, 2, 3$ and velocity coordinates $v \in \Omega_v \subset \mathbb{R}^d$. A grid-based discretization with n grid points in each dimension therefore incurs a computational cost of order $O(n^{2d})$. Hence, direct solutions for the (kinetic) density function are prohibitively expensive.

Recently, dynamical low-rank solvers for kinetic equations such as the Vlasov-Poisson equation have been proposed [4]. These methods are based on [7], in which matrix differential equations are approximated by low-rank matrices using a Galerkin principle. These solvers have been shown to be successful [1], but without additional care the projection to low-rank tensor products destroys important physical structure in the system. This can be alleviated by altering the approximation space such that some moments of the velocity coordinate always lie in the space spanned by the low-rank matrices [5]. This is sufficient to guarantee that continuity equations for mass, momentum, and energy are conserved up to an error that is linear in the time step size.

In the following report we examine whether the conservative dynamical low-rank integrator also preserves further observables and if so, to what order. The report is structured as follows: in section 2 the Vlasov-Poisson equation is introduced and conservation properties are discussed; in section 3 the conservative dynamical low-rank integrator is shown; in section 4.1 the integrator is demonstrated on an analytically understood example system.

if some observable works, mention here

2 The Vlasov-Poisson Equation and Conservation Properties

The Vlasov-Poisson equation models the evolution of the electron density $f = f(x, v, t)$ in plasma. It is assumed that (1) slow movement of heavy positive ions, (2) pair collisions between electrons, (3) variations in magnetic field, and (4) relativistic effects are all negligible. The resultant (non-dimensionalized) evolution equation reads

$$\partial_t f(x,v,t) + v \cdot \nabla_x f(x,v,t) + E(x,t) \cdot \nabla_v f(x,v,t) = 0 \tag{2.1}$$

where $E = E(f)$ is the self-consistent electric field. The mass

$$\mathcal{M}_0(t) = \int_{\Omega_x} \int_{\Omega_v} f(x,v,t)\; dv\, dx \tag{2.2}$$

is wlog. assumed to be initially equal to 1. Moreover E is written as a potential $E = -\nabla\phi$ where

$$-\Delta\phi\,(x) = 1 - \int_{\Omega_v} f(x,v,t)\; dv. \tag{2.3}$$

Equation 2.1 contains infinitely many invariants. Indeed, all L^p norms $\|f\|_p$ as well as entropy

$$\mathcal{S}(t) = \iint_{\Omega_x \times \Omega_v} f(x,v,t)\; \log f(x,v,t)\; dv\, dx \tag{2.4}$$

are conserved [**<empty citation>**]. Additionally it can be shown that all moments of the velocity distrbution

cite entropy conservation

$$M_k(x,t) = \int_{\Omega_v} vv \dots v f(x,v,t)\; dv \tag{2.5}$$

(wich are k-tensors) satisfy continuity equations and correspondingly their space-integrals $\mathcal{M}_k = \int_{\Omega_x} M_k\, dx$ are also conserved. Specifically, for a cartesian index $\iota = (\iota_1, \dots, \iota_k) \in \{1, \dots, d\}^k$ of M_k there is a multiindex $\varrho = (\varrho_1, \dots \varrho_d) \in \mathbb{N}^d$ with $|\iota| = k$ and

$$M_k^\iota = \int_{\Omega_v} v_{\iota_1} v_{\iota_2} \dots v_{\iota_k}\, f\, dv = \int_{\Omega_v} v_1^{\varrho_1} v_2^{\varrho_2} \dots v_d^{\varrho_d}\, f\, dv. \tag{2.6}$$

The continuity equation for M_k^ι is written

$$\partial_t \int_{\Omega_v} v^\varrho f \, dv + \nabla_x \cdot \int_{\Omega_v} v^{\varrho+1} f \, dv + E \cdot \int_{\Omega_v} \varrho v^{\varrho-1} f \, dv = 0 \tag{2.7}$$

where $z^{\varrho \pm 1}$ is understood as the vector

$$z^{\varrho\pm1} = \begin{bmatrix} z^{(\varrho_1 \pm 1,\, \varrho_2,\, \dots,\, \varrho_d)} \\ z^{(\varrho_1,\, \varrho_2 \pm 1,\, \dots,\, \varrho_d)} \\ \vdots \\ z^{(\varrho_1,\, \dots,\, \varrho_{d-1},\, \varrho_d \pm 1)} \end{bmatrix}. \tag{2.8}$$

Notable are the first few moments and their continuity equations

$$\partial_t M_0 + \nabla_x \cdot M_1 = 0, \tag{2.9}$$
$$\partial_t M_1 + \nabla_x \cdot M_2 + E M_0 = 0, \tag{2.10}$$

which imply conservation of mass $\mathcal{M}_0$ and momentum $\mathcal{M}_1$ after integrating over Ω_x. The continuity equation for (total) energy density is deduced from the trace of the second moment

$$\partial_t e + \frac{1}{2} \nabla_x \cdot (\mathrm{tr} M_3) = (\partial_t E - M_1) \cdot E \tag{2.11}$$

where

$$e(x,t) = \frac{1}{2} \mathrm{tr} M_2(x,t) + \frac{1}{2} E(x,t)^2 = \frac{1}{2} \int_{\Omega_v} v^2 f(x,v,t) \ dv + \frac{1}{2} E(x,t)^2 \tag{2.12}$$

and

$$\mathrm{tr} M_3(x,t) = \int_{\Omega_v} v^2 \ v f(x,v,t) \ dv. \tag{2.13}$$

this is trace of a 3-tensor ie. tensor contraction - should maybe simplify?

The dynamical low-rank integrator presented in the following section is designed to satisfy equations 2.9, 2.10, and 2.11.

3 A Low-rank Tensor Approximation Scheme

3.1 Presentation of the Integrator

The density function $f(x,v,t) \in L^2(\Omega_x \times \Omega_v)$ is approximated by a tensor product of functions in x and in v:

$$f(x,v,t) = f_0(v) \sum_{i,j=1}^{r} X_i(x,t) \ S_{ij}(t) \ V_j(v,t) =: f_0(v) \ X(x,t)^T \ S(t) \, V(v,t) \tag{3.1}$$

where r is the approximation rank and $f_0(v) = \exp(-v^2)$ is a Gaussian weight.

Let

$$\langle g, h\rangle_x = \int_{\Omega_x} f\, g\ dx, \quad (g, h)_v = \int_{\Omega_v} g\, h\ dv \quad \langle g, h\rangle_v = \int_{\Omega_v} g\, h\, f_0\ dv. \tag{3.2}$$

We require that X and V satisfy the orthonormality conditions

$$\langle X_i,\, X_j\rangle_x = \delta_{ij}, \quad \langle V_i,\, V_j\rangle_v = \delta_{ij}, \quad 1 \le i, j \le r \tag{3.3}$$

where δ_{ij} refers the the Kronecker delta, as well as the gauge conditions

$$\langle \partial_t X_i,\, X_j\rangle_x = 0, \quad \langle \partial_t V_i,\, V_j\rangle_v = 0, \quad 1 \le i, j \le r. \tag{3.4}$$

Let further $\bar{X} = \operatorname{span}\{X_1, \ldots, X_r\}$, $\bar{V} = \operatorname{span}\{V_1, \ldots, V_r\}$. The Galerkin condition yields the equation [4]:

$$\partial_t f = \Pi\,(\,RHS(f)\,) \quad \text{where} \quad \Pi\, g = \Pi_{\bar{V}}\, g - \Pi_{\bar{X}}\Pi_{\bar{V}}\, g + \Pi_{\bar{X}}\, g \tag{3.5}$$

and $\Pi_{\bar{X}}$, $\Pi_{\bar{V}}$ are the orthogonal projections onto $\bar{X}$ and $\bar{V}$, respectively. A first-order Lie-Trotter splitting based on the three terms in equation 3.5 yields equations of motion for the components of X, S, and V.

The key insight of [3] is that if the functions $v \mapsto 1$, $v \mapsto v_1, \ldots, v \mapsto v_d$, and $v \mapsto v^2$ lie in $\bar{V}$, then discrete versions of equations 2.9, 2.10, and 2.11 hold. Hence the integration scheme is altered to guarantee this condition. We split $V(v,t) \in \mathbb{R}^r$ into two blocks $U(v,t) \in \mathbb{R}^m$ and $W \in \mathbb{R}^{r-m}$:

$$V = \begin{bmatrix} U \\ W \end{bmatrix} \tag{3.6}$$

where U is fixed throughout the integration and contains the desired functions. To satisfy the orthonormality condition 3.3 we use Hermite polynomials

$$U(v,t) = U(v) \propto \begin{bmatrix} 1 \\ 2v_1 \\ \vdots \\ 2v_d \\ 4v^2 - 2 \end{bmatrix}. \tag{3.7}$$

Performing the analogous calculations as in [4] for the altered basis functions yields [3]

$$\sum_i \partial_t X_i \, S_{ik} = (\, V_k,\, RHS(f)\,)_v - \sum_i X_i \, \partial_t S_{ik}, \qquad 1 \le k \le r \tag{3.8}$$

$$\sum_{ip} S_{i\,q+m} \, S_{ip} \, \partial_t W_p = \frac{1}{f_0} \sum_i S_{i\,q+m} \, \langle\, X_i,\, RHS(f)\,\rangle_x - \sum_{il} S_{i\,q+m} \, \partial_t S_{il} \, V_l, \qquad 1 \le q \le r-m \tag{3.9}$$

$$\partial_t S_{kl} = (\, X_k,\, (V_l,\, RHS(f))_v\,)_x \qquad 1 \le k,l \le r. \tag{3.10}$$

Concrete equations for computing the above inner products are given in [5]. It should be reemphasized that the orthonormality and guage conditions 3.3, 3.4 must still hold, so the components of U must be appropriately scaled.

To solve equations 3.8, 3.9 directly, the matrix S must be inverted. However, as the approximation rank r increases, S has progressively smaller singular values. Hence the scheme becomes increasingly ill-conditioned as the accuracy increases. We therefore need to alter the low-rank scheme again to address this issue.

Notice that the low-rank approximation $f = X^T SV$ can be written as $f = K^T V$ for some $K = K(x,t) \in \mathbb{R}^r$. X^T and S are then (up to a unitary basis transformation) the result of a (semidiscrete[1]) QR factorization of K^T. Analogously, f can be written as $f = X^T L$ and the elements W can also be reconstructed by a QR factorization. We may therefore rewrite equations 3.8 and 3.9 as

$$\partial_t K_k = (\, V_k,\, RHS(f)\,)_v, \qquad 1 \le k \le r \tag{3.11}$$

$$\partial_t L_q = \frac{1}{f_0} \sum_i S_{i\,q+m} \, \langle\, X_i,\, RHS(f)\,\rangle_x - \sum_{il} S_{i\,q+m} \, \partial_t S_{il} \, V_l, \qquad 1 \le q \le r-m \tag{3.12}$$

mMn sollte die Ableitung im letzten Term eigentlich aud das erste S sein, aber so haben die es im Einkemmer Paper gemacht und es functioniert

with

$$K_k = \sum_i X_i \, S_{ik}, \qquad L_q = \sum_{il} S_{i\,q+m} \, S_{il} W_l. \tag{3.13}$$

Using a stepping scheme

$$K(t+\tau) = K(t) + \tau \partial_t K(t), \qquad L(t+\tau) = L(t) + \tau \partial_t L(t) \tag{3.14}$$

we may obtain $X(t+\tau)$, $V(t+\tau)$ via QR factorization

$$K_k(t+\tau) = \sum_i X_i(t+\tau) \, R^1_{ik}, \qquad L_q(t+\tau) = \sum_i W_i(t+\tau) \, R^2_{iq}. \tag{3.15}$$

Finally, we compute $S(t+\tau)$ using equation 3.10 and a best-approximation of f:

[1] The standard Gram-Schmidt process to construct a QR decomposition can just as easily be viewed in a semidiscrete setting: each "column" of K^T is a function of x evaluated across Ω_x, instead of each column being a discrete vector. The algorithm does not need to be changed at all.

$$f \approx \widetilde{f} := f_0 \, X(t+\tau)^T \, (M^T S N) \, V(t+\tau), \tag{3.16}$$

$$S_{kl}(t+\tau) = \sum_{ij} M^T_{ki} \, S_{ij} \, N_{jl} + \tau \left(X_k, \, \left(V_l, \, RHS(\widetilde{f}) \right)_v \right)_x \tag{3.17}$$

where

$$M_{ki} = \langle X_k(t), \, X_i(t+\tau) \rangle_x , \qquad N_{jl} = \langle V_j(t), \, V_l(t+\tau) \rangle_v . \tag{3.18}$$

Rename M and N so that it doesn't conflict with moments

Crucially, the approximation $f \approx \widetilde{f}$ doe *not* conserve any of the invariants since the projections in 3.18 are not conservative. Therefore we need to expand the basis onto which we project.

Specifically, let $(\widetilde{X}_j)_j$ be an orthonormal basis of span $\{X_i(t), \nabla X_i(t), E(t)X_i(t), K_i(t+\tau)\}_i$ and $(\widetilde{V}_j)_j$ an orthonormal basis of span $\{V_i(t), L_q(t+\tau)\}_{iq}$. *We emphasize that this definition of* $\widetilde{X}$ *differs slightly from previous works due to the addition of* EX. This drastically reduces the complexity of future proofs, while only increasing the basis size linearly.

The projections

$$\widetilde{M}_{ki} = \left\langle \widetilde{X}_k, \, X_i(t) \right\rangle_x , \qquad \widetilde{N}_{jl} = \left\langle \widetilde{V}_j, \, V_l(t) \right\rangle_v \tag{3.19}$$

are mass, momentum, and energy conservative [5]. However, this has increased the rank of the approximation f. Thus, we need to truncate the approximation in a way which ensures that the fixed basis functions of U remain unchanged. For convenience write

$$\widetilde{S}(t) = \widetilde{M}^T S(t) \widetilde{N} \tag{3.20}$$

where M and N are as in equation 3.19, and write $\widetilde{S}(t+\tau)$ as the result of applying equation 3.17 with $\widetilde{M}$, $\widetilde{S}(t)$, $\widetilde{N}$, $\widetilde{X}$, and $\widetilde{V}$. Letting $\widetilde{K}^T = \widetilde{X}^T \widetilde{S}(t+\tau)$ and using the structure of $\widetilde{V}$,

$$f(t+\tau) \approx \widetilde{K}^T \widetilde{V} = \begin{bmatrix} (\widetilde{K}^{cons})^T & (\widetilde{K}^{rem})^T \end{bmatrix} \begin{bmatrix} U \\ \widetilde{W} \end{bmatrix} \tag{3.21}$$

where $\widetilde{K}^{cons}$ is the first m components of $\widetilde{K}$, and $\widetilde{K}^{rem}$, $\widetilde{W}$ are the last components of $\widetilde{K}$, $\widetilde{V}$ respectively.

Hence, by truncating $\widetilde{K}^{rem}$ and $\widetilde{W}$, the desired components of U remain unaffected. We perform the truncation as follows: QR factorizations of $\widetilde{K}^{cons}$, $\widetilde{K}^{rem}$ yield

$$\widetilde{K}^{cons}_k = \sum_i X^{cons}_i \, S^{cons}_{ik} , \qquad \widetilde{K}^{rem}_q = \sum_j \widetilde{X}^{rem}_j \, \widetilde{S}^{rem}_{jq} . \tag{3.22}$$

By a truncated singular value decomposition of $\widetilde{S}^{rem}$, keeping only the largest $r - m$ singular values, we have

$$\widetilde{S}^{rem} \approx \widehat{U}\widehat{S}\widehat{W}. \tag{3.23}$$

Now set $S^{rem} = \widehat{S}$ and

$$X_q^{rem} = \sum_i \widetilde{X}_i^{rem}\,\widehat{U}_{iq}, \qquad \breve{W}_q = \sum_j \widetilde{W}_j\,\widehat{W}_{jq}, \qquad 1 \leq q \leq r-m. \tag{3.24}$$

Combining $\widehat{X} = \begin{bmatrix} X^{cons} \\ X^{rem} \end{bmatrix}$ and performing a final QR factorization

$$\widehat{X}_k = \sum_i \check{X}_i\,R_{ik} \tag{3.25}$$

finishes the truncation, as we set

$$X(t+\tau) = \check{X}, \qquad S(t+\tau) = R\begin{bmatrix} S^{cons} & \\ & S^{rem} \end{bmatrix}, \qquad V(t+\tau) = \begin{bmatrix} U \\ \breve{W} \end{bmatrix}. \tag{3.26}$$

While in this case we have performed the time-stepping in equations 3.21, 3.17 via a simple explicit Euler scheme, the extension to time steps of higher order is immediate. Indeed, we refer to [2] for an extension of the robust integrator using the midpoint rule. Pseudocode for the presented algorithm can be found in [5] and implementation of the algorithm (as well as the midpoint-rule extension) in [6].

3.2 Discrete Conservation Equations

We show that the integrator proposed satisfies the moment continuity equations 2.9 - 2.11 as well as preserves the electric energy up to first oder.

Lemma 3.1. *Let $\left(\widetilde{X}_k\right)_k$ be constructed as in 3.19. Then $\left(\widetilde{X}_k\right)_k$ spans $RHS(\widetilde{f})$, that is,*

$$\sum_k \widetilde{X}_k \left\langle \widetilde{X}_k,\ RHS(\widetilde{f}) \right\rangle_x = RHS(\widetilde{f}). \tag{3.27}$$

Proof. We note first that due to the augmentation of the basis, $\widetilde{f} = f = f_0 X^T S V$. The claim now follows from a direct calculation using the fact that $\widetilde{X}$ spans a basis of $\nabla_x X$ and EX

$$\begin{aligned}
\sum_k \widetilde{X}_k \left\langle \widetilde{X}_k,\ RHS(\widetilde{f}) \right\rangle_x &= \sum_{ij} S_{ij}\left(\sum_k \widetilde{X}_k \left\langle \widetilde{X}_k,\ \nabla_x X_i \right\rangle_x\right) \cdot vV_j \\
&\quad + \sum_{ij} S_{ij}\left(\sum_k \widetilde{X}_k \left\langle \widetilde{X}_k,\ EX_i \right\rangle_x\right) \cdot \nabla_v\left[f_0 V_j\right] \\
&= \sum_{ij} S_{ij} \nabla_x X_i \cdot vV_j + \sum_{ij} S_{ij} EX_i \cdot \nabla_v\left[f_0 V_j\right] = RHS(f).
\end{aligned} \tag{3.28}$$

□

Lemma 3.2. *Let $\widetilde{K}$ be constructed as in 3.21. Then for $\ell \leq m$,*

$$\left(U_\ell, f(t+\tau)\right)_v = X(t+\tau)^T S(t+\tau) \left\langle U_\ell, V(t+\tau)\right\rangle_v = \widetilde{K}_\ell. \tag{3.29}$$

The above lemma effectively states that the truncation step is indeed conservative. This is crucial in the proceeding theorem.

Proof. Calculate

$$\begin{aligned}\left(U_\ell,\ f(t+\tau)\right)_v &= \sum_{ij} X_i(t+\tau) S_{ij}(t+\tau) \left\langle U_\ell,\ V_j\right\rangle_v \\ &= \sum_i X_i(t+\tau) S_{i\ell}(t+\tau) \\ &= \sum_i X_i(t+\tau) \sum_j R_{ij} \begin{bmatrix} S^{cons} & \\ & S^{rem} \end{bmatrix}_{j\ell}\end{aligned} \quad . \tag{3.30}$$

Since $\ell \leq m$ (recall V_ℓ for $\ell > m$ was denoted $W_{\ell-m}$) and $S^{cons} \in \mathbb{R}^{m\times m}$,

$$\begin{aligned}\sum_i X_i(t+\tau) \sum_j R_{ij} \begin{bmatrix} S^{cons} & \\ & S^{rem} \end{bmatrix}_{j\ell} &= \sum_{j=1}^m \sum_i X_i(t+\tau) R_{ij} S^{cons}_{j\ell} \\ &= \sum_{j=1}^m \widehat{X}_j S^{cons}_{j\ell} \\ &= \sum_i X^{cons}_i S^{cons}_{i\ell} \\ &= \widetilde{K}_\ell.\end{aligned} \tag{3.31}$$

□

Theorem 3.3. *Consider the P-th order moment M^ι_P with corresponding multiindex ϱ as in equation 2.6. M^ι_P satisfies the time-discrete continuity equation*

$$\begin{aligned}\frac{M^\iota_{P+1}(t+\tau) - M^\iota_{P+1}(t)}{\tau} + \nabla_x \cdot \int_{\Omega_v} v^{\varrho+1} f(t)\,dv + E(t) \cdot \int_{\Omega_v} \varrho v^{\varrho-1} f(t)\,dv \\ = \left(\Pi_{(spanU)^\perp}\left[v^\varrho\right],\ \frac{f(t+\tau)-f(t)}{\tau} - RHS(f(t))\right)_v\end{aligned} \tag{3.32}$$

where

$$\Pi_{(spanU)^\perp}[g] = g - \sum_\ell U_\ell \left\langle U_\ell,\ g\right\rangle_v. \tag{3.33}$$

Remark. 1. In particular, when $U : \Omega_v \to \mathbb{R}^m$ is such that

$$v \mapsto v^{\varrho} \in \operatorname{span}\{U_1, \dots, U_m\} \tag{3.34}$$

then the right hand side of equation 3.32 is zero. Hence, when equation 3.34 holds for all multiindices with $|\varrho| = P$, then the continuity equation for the tensor M_P holds.

2. The error representation in the form of an inner product is much smaller than the norm-based error bounds typically given in numerical analysis. Indeed, consider the scalar-valued error estimate

$$\begin{aligned} &\left\| \frac{M^{\iota}_{P+1}(t+\tau) - M^{\iota}_{P+1}(t)}{\tau} + \nabla_x \cdot \int_{\Omega_v} v^{\varrho+1} f(t)\, dv + E(t) \cdot \int_{\Omega_v} \varrho v^{\varrho-1} f(t)\, dv \right\|_x^2 \\ &= \left\| \left\langle \Pi_{(\operatorname{span} U)^\perp}\left[v^{\varrho}\right],\ \frac{f(t+\tau) - f(t)}{f_0 \tau} - \frac{1}{f_0} RHS(f(t)) \right\rangle_v \right\|_x^2 \\ &\leq \left\| \Pi_{(\operatorname{span} U)^\perp}\left[v^{\varrho}\right] \right\|_v \int_{\Omega_x} \left\| \frac{f(t+\tau) - f(t)}{f_0 \tau} - \frac{1}{f_0} RHS(f(t)) \right\|_v dx. \end{aligned} \tag{3.35}$$

where for the inequality we used the Cauchy-Schwarz inequality. Indeed, the inequality is highly pessimistic, the latter term is often multiple orders of magnitude larger than the former, see section 4.1.

Proof. We first note that the discrete continuity equation can be written as

$$\left(v^{\varrho},\ \frac{f(t+\tau) - f(t)}{\tau} - RHS(f(t)) \right)_v. \tag{3.36}$$

Hence, the claim of the theorem is that when $v \mapsto v^{\varrho} \in \operatorname{span}\{U_1, \dots, U_m\}$, the expression 3.36 is zero. To that end, split v^{ϱ} into two components

$$v^{\varrho} = \sum_{\ell} \xi_\ell U_\ell(v) + \left(\Pi_{(\operatorname{span} U)^\perp}\left[v^{\varrho}\right] \right)(v). \tag{3.37}$$

We begin by giving a representation of 3.36 for the component of v^{ϱ} in the span of U. Recalling that $\widetilde{K}_\ell = \sum_k \widetilde{X}_k \widetilde{S}_{k\ell}(t+\tau)$ and $\widetilde{S}_{k\ell}(t+\tau) = \widetilde{S}_{k\ell} + \tau \left(\widetilde{X}_k U_\ell,\ RHS(\widetilde{f}) \right)_{xv}$, we may write

$$\begin{aligned} \widetilde{K}_\ell &= \sum_k \widetilde{X}_k \left[\sum_{ij} \left\langle \widetilde{X}_k,\ X_i(t) \right\rangle_x S_{ij}(t) \left\langle V_\ell(t),\ \widetilde{V}_j \right\rangle_v + \tau \left(\widetilde{X}_k U_\ell,\ RHS(\widetilde{f}) \right)_{xv} \right] \\ &= \sum_i S_{i\ell}(t) \sum_k \widetilde{X}_k \left\langle \widetilde{X}_k,\ X_i(t) \right\rangle_x + \tau \sum_k \widetilde{X}_k \left(\widetilde{X}_k U_\ell,\ RHS(\widetilde{f}) \right)_{xv}. \end{aligned} \tag{3.38}$$

The first term in equation 3.38 is equal to

$$\sum_i S_{i\ell}(t) \sum_k \widetilde{X}_k \left\langle \widetilde{X}_k,\ X_i(t) \right\rangle_x = \sum_i S_{i\ell}(t) X_i(t) = K_\ell(t) = (U_\ell,\ f(t))_v\,. \tag{3.39}$$

Hence,

$$\left(U_\ell,\ \frac{f(t+\tau) - f(t)}{\tau} \right)_v = \sum_k \widetilde{X}_k \left(\widetilde{X}_k U_\ell,\ RHS(\widetilde{f}) \right)_{xv}. \tag{3.40}$$

Using the splitting in equation 3.37,

$$\left(v^\varrho,\ \frac{f(t+\tau) - f(t)}{\tau} \right)_v = \left(\sum_\ell \xi_\ell U_\ell,\ \frac{f(t+\tau) - f(t)}{\tau} \right)_v + \left(\Pi_{(\mathrm{span} U)^\perp} \left[v^\varrho \right],\ \frac{f(t+\tau) - f(t)}{\tau} \right)_v. \tag{3.41}$$

The first term is rewritten to

$$\begin{aligned}
\left(\sum_\ell \xi_\ell U_\ell,\ \frac{f(t+\tau) - f(t)}{\tau} \right)_v &= \sum_k \widetilde{X}_k \left(\widetilde{X}_k \sum_\ell \xi_\ell U_\ell,\ RHS(\widetilde{f}) \right)_{xv} \\
&= \sum_k \widetilde{X}_k \left(\widetilde{X}_k v^\varrho,\ RHS(\widetilde{f}) \right)_{xv} \\
&\qquad - \sum_k \widetilde{X}_k \left(\widetilde{X}_k \Pi_{(\mathrm{span} U)^\perp} \left[v^\varrho \right],\ RHS(\widetilde{f}) \right)_{xv} \\
&= \left(v^\varrho - \Pi_{(\mathrm{span} U)^\perp} \left[v^\varrho \right],\ \sum_k \widetilde{X}_k \left\langle \widetilde{X}_k,\ RHS(\widetilde{f}) \right\rangle_x \right)_v \\
&= \left(v^\varrho - \Pi_{(\mathrm{span} U)^\perp} \left[v^\varrho \right],\ RHS(f) \right)_v
\end{aligned} \tag{3.42}$$

where for the last equality we have used lemma 3.1. Inserting equation 3.42 into equation 3.41 and rearranging yields the claim.

□

Due to the orthonormality condition on U, the most reasonable choice to satisfy theorem 3.3 is given by Hermite polynomials, that is, we choose tensor products of Hermite polynomials up to degree P. Setting $P = 0$ implies conservation of mass and $P = 1$ implies conservation of momentum. To see conservation of total energy, a final calculation must be made.

Corollary 3.4. *Let $U : \Omega_v \to \mathbb{R}^m$ be chosen such that*

$$v \mapsto v^2 \in \mathit{span}\{U_1, \ldots, U_m\} \tag{3.43}$$

Assume that the numerically computed electric field satisfies

$$\|E(t+\tau) - E(t)\| = \mathcal{O}(\tau^\mu) \tag{3.44}$$

for some $\mu \geq 1$. Then the time-discrete continuity equation for total energy is satisfied to order $2\mu - 1$, that is,

$$\frac{e(t+\tau) - e(t)}{\tau} + \frac{1}{2}\nabla_x \cdot (trM_3(t)) + (M_1(t) - \frac{E(t+\tau) - E(t)}{\tau}) \cdot E(t) = \mathcal{O}(\tau^{2\mu-1}). \tag{3.45}$$

Proof. Note that

$$\frac{e(t+\tau) - e(t)}{\tau} = \frac{1}{2}\frac{\mathrm{tr}M_2(t+\tau) - \mathrm{tr}M_2(t)}{\tau} + \frac{1}{2}\frac{E(t+\tau)^2 - E(t)^2}{\tau}, \tag{3.46}$$

the first term of which is equal to

$$\frac{\mathrm{tr}M_2(t+\tau) - \mathrm{tr}M_2(t)}{\tau} = -\nabla_x \cdot (\mathrm{tr}M_3(t)) - 2M_1(t) \cdot E(t) \tag{3.47}$$

by theorem 3.3. The second term can be rewritten as

$$\begin{aligned}\frac{E(t+\tau)^2 - E(t)^2}{\tau} &= \frac{(E(t+\tau) - E(t))^2 + 2E(t+\tau) \cdot E(t) - 2E(t)^2}{\tau} \\ &= 2E(t) \cdot \frac{E(t+\tau) - E(t)}{\tau} + \frac{(E(t+\tau) - E(t))^2}{\tau}.\end{aligned} \tag{3.48}$$

Since $E(t+\tau)$ is always an $\mathcal{O}(\tau^\mu)$ away from $E(t)$, the second term is an $\mathcal{O}(\tau^{2\mu-1})$. Inserting the expressions 3.47 and 3.48 into equation 3.46 now yields the claim. ☐

Theorem 3.3 states that when all monomials up to order P are held constant in U, then the corresponding continuity equations up to order P are satisfied. For higher order continuity equations, we can similarly guarantee a worst-case bound on the error, dependent on how well the discrete differential $\frac{1}{\tau}(f(t+\tau) - f(t))$ matches $RHS(f)$ in the direction of some monomial $v \mapsto v^\varrho$.

4 Numerical Examples

We demonstrate the error in observables for multiple initial conditions. We show that the proposed methods indeed satisfy conservation laws as stated in theorem 3.3, as well as showing that the error representation is significantly smaller than the worst-case estimate given in equation 3.35.

4.1 (Non-)Linear Landau Damping

A key qualitative difference in the Vlasov equation compared to fluid models is the ability for dissipative behavior to occur without the loss of energy. We are able to see this empirically, as was verified analytically in [8] using asymptotic methods for small parameter values. We consider a perturbation of the trivial steady state distribution

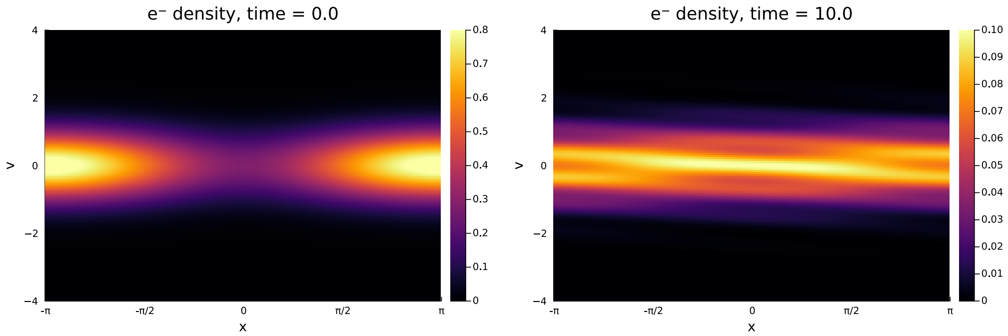

Figure 4.1: Landau damping. Density **??** and RHS for various times. !! Note the differing color bar axes!

$$f(x, v, 0) = \frac{1}{2\pi}(1 + \alpha\cos(x))\exp(-v^2) \tag{4.1}$$

for $\alpha \in [0, 1/2]$, computed over the domain $\Omega_x = [-\pi, \pi)_{\text{periodic}}$, $\Omega_v = [-6, 6]$. We use a grid of 512 equally spaced points in Ω_x, 512 Gauß-Legendre grid points transplanted to Ω_v. Furthermore we fix a rank $r = 10$ and a time step $\tau = 5 \cdot 10^{-4}$.

4.2 Two-stream Instability

The initial conditions in the Landau damping case can often show positive results due to the high amount of symmetry in the system. For this reason, we consider a nonsymmetric two-stream instabilility

$$f(x, v, 0) = \frac{1}{4\pi}(1 + \alpha\cos(x))\exp(-(v - \bar{v})^2) + \frac{1}{4\pi}(1 + \beta\sin(x))\exp(-(v + \bar{v})^2) \tag{4.2}$$

for $\bar{v} \in \mathbb{R}$ and $\alpha, \beta \in [0, 1/2]$, chosen such that $f(0)$ is strictly positive. We use the same grid points, rank, and time step as in the previous section.

5 Conclusion

References

[1] Fabio Cassini and Lukas Einkemmer. "Efficient 6D Vlasov simulation using the dynamical low-rank framework Ensign". In: *Computer Physics Communications* 280 (2022), p. 108489. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2022.108489. URL: http://dx.doi.org/10.1016/j.cpc.2022.108489.

[2] Gianluca Ceruti et al. *A robust second-order low-rank BUG integrator based on the midpoint rule*. 2024. arXiv: 2402.08607 [math.NA]. URL: https://arxiv.org/abs/2402.08607.

[3] Lukas Einkemmer and Ilon Joseph. "A mass, momentum, and energy conservative dynamical low-rank scheme for the Vlasov equation". In: *Journal of Computational Physics* 443 (2021), p. 110495. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2021.110495. URL: http://dx.doi.org/10.1016/j.jcp.2021.110495.

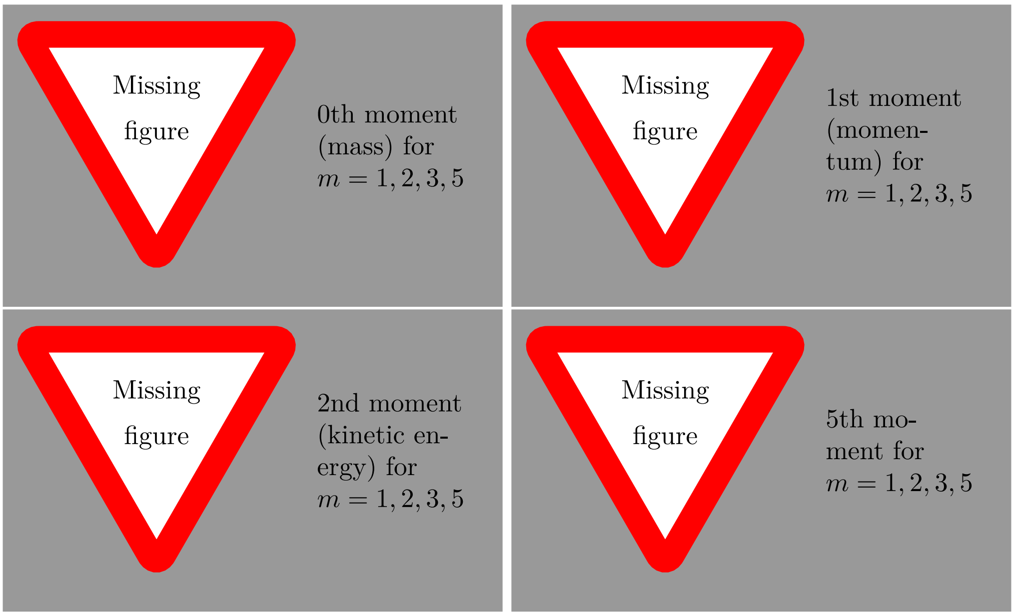

Figure 4.2: Error in the continuity equations for the Landau damping scenario with $\alpha = 10^{-3}$.

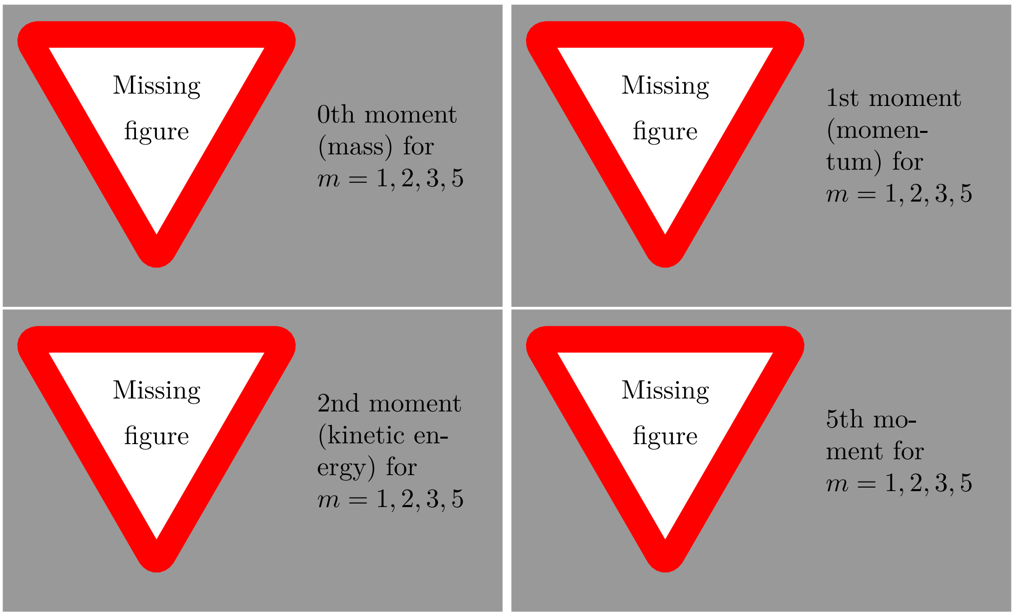

Figure 4.3: Error in the continuity equations for the Landau damping scenario with $\alpha = 1/2$.

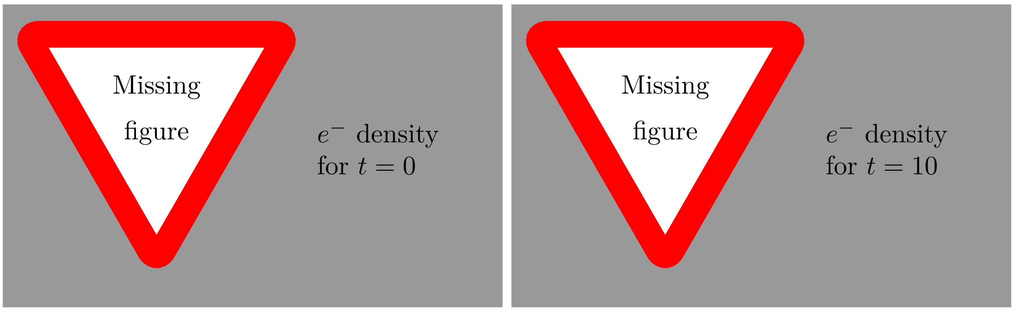

Figure 4.4: Two-stream instability. density for various times.

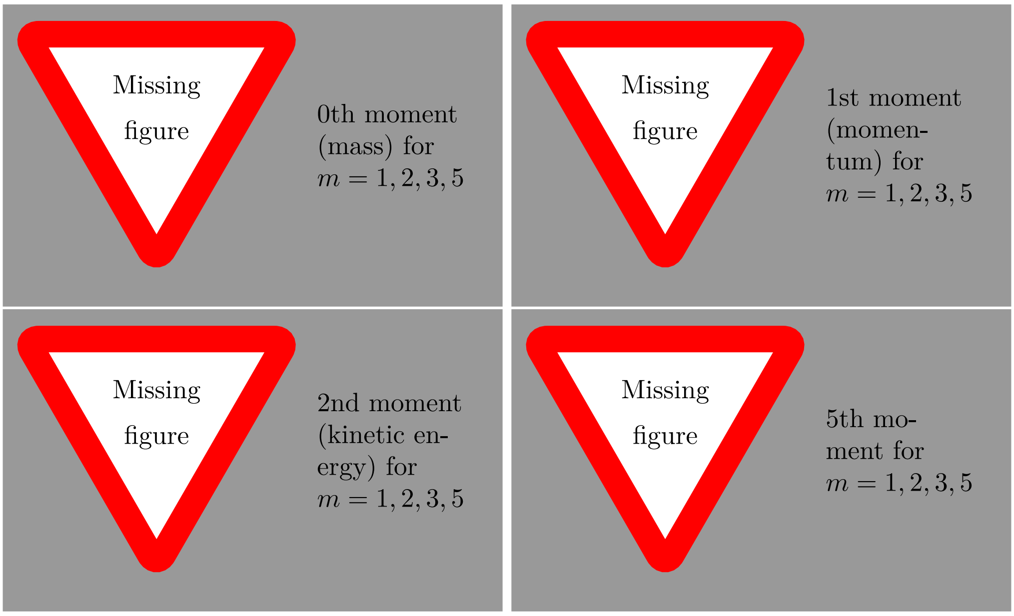

Figure 4.5: Error in the continuity equations for the two-stream instability.

[4] Lukas Einkemmer and Christian Lubich. *A low-rank projector-splitting integrator for the Vlasov–Poisson equation.* 2018. eprint: 1801.01103. URL: https://arxiv.org/abs/1801.01103.

[5] Lukas Einkemmer, Alexander Ostermann, and Carmela Scalone. “A robust and conservative dynamical low-rank algorithm”. In: *Journal of Computational Physics* 484 (2023), p. 112060. ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2023.112060.

[6] April Herwig. *DynamicalLowRankIntegrators.jl.* https://github.com/April-Hannah-Lena/DynamicalLowRankIntegrators.jl. 2025.

[7] Othmar Koch and Christian Lubich. “Dynamical Low-Rank Approximation”. In: *SIAM Journal on Matrix Analysis and Applications* 29.2 (2007), pp. 434–454. DOI: 10.1137/050639703. URL: https://doi.org/10.1137/050639703.

[8] Lev Davidovich Landau. “On the vibrations of the electronic plasma”. In: *Yad. Fiz.* 10 (1946), p. 25. URL: https://cds.cern.ch/record/437300.