

TRANSPORT BASED PARTICLE METHODS FOR THE FOKKER-PLANCK-LANDAU EQUATION*

VASILY ILIN[†], JINGWEI HU[‡], AND ZHENFU WANG[§]

Abstract. We propose a particle method for numerically solving the Landau equation, inspired by the score-based transport modeling (SBTM) method for the Fokker-Planck equation. The semi-discretized scheme can preserve some important physical properties of the Landau equation, such as the conservation of mass, momentum, and energy, and decay of estimated entropy. We prove that matching the gradient of the logarithm of the approximate solution is enough to recover the true solution to the Landau equation with Maxwellian molecules. Several numerical experiments in low and moderately high dimensions are performed, with particular emphasis on comparing the proposed method with the traditional particle or blob method.

Keywords. Fokker-Planck-Landau equation; Maxwellian molecule; particle method; Kullback–Leibler divergence; neural network; score matching.

AMS subject classifications. 35Q84; 65M75; 49Q22; 68T07.

1. Introduction

The Vlasov-Landau equation [24] is a kinetic equation that describes the evolution of the probability density $f(t, x, v)$ of a plasma, with position $x \in \Omega \subset \mathbb{R}^d$ and velocity $v \in \mathbb{R}^d$, $d \geq 2$:

$$\begin{aligned} \frac{\partial f}{\partial t} + v \cdot \nabla_x f + (E + v \times B) \cdot \nabla_v f &= Q(f, f), \\ Q(f, f)(t, x, v) &= \nabla_v \cdot \int_{\mathbb{R}^d} A(v-w) [f(t, x, w) \nabla_v f(t, x, v) - f(t, x, v) \nabla_w f(t, x, w)] dw, \end{aligned} \tag{1.1}$$

where E and B are the electric and magnetic fields given externally or determined self-consistently by the Maxwell’s equations. The integro-differential nonlinear operator $Q(f, f)$ is the so-called Landau operator which models collisions between charged particles. It can be derived from the Boltzmann collision operator by taking the scattering angle of collisions to be small. The kernel A is a $d \times d$ matrix given by $A_{i,j}(z) = |z|^\gamma (|z|^2 \delta_{i,j} - z_i z_j)$. The parameter $-d-1 \leq \gamma \leq 1$ determines the strength of particle interactions. The most physically relevant case is $\gamma = -d = -3$, which corresponds to Coulomb interactions. The case $\gamma = 0$ corresponds to the so-called Maxwellian molecules, which admits many mathematical simplifications and hence is also widely considered in analysis and computation.

In this work, we will neglect spatial dependence in Equation (1) and focus on the spatially homogeneous Landau equation:

$$\frac{\partial f}{\partial t} = Q(f, f), \tag{1.2}$$

where $f = f(t, v)$ is a function of t and v only.

*Received: June 03, 2024; Accepted (in revised form): February 15, 2025. Communicated by Lorenzo Pareschi.

[†]Department of Mathematics, University of Washington, Seattle, WA, USA (vilin@uw.edu).

[‡]Department of Applied Mathematics, University of Washington, Seattle, WA, USA (hujw@uw.edu).

[§]Beijing International Center for Mathematical Research, Peking University, Beijing, China (zwang@bicmr.pku.edu.cn).

A comprehensive study of the Landau equation with Maxwellian molecules was done by Villani in [26, 23]. The main results are existence and uniqueness of smooth solutions, conservation of mass, momentum and energy, and decay of entropy and Fisher information. The author also showed how to rewrite the Landau equation in the form of the Fokker-Planck equation with a time-and-space-dependent diffusion matrix. This motivates applying the tools, which are used for the Fokker-Planck equation, to the Landau equation. Also due to this analogy, the Landau equation is often referred to as the Fokker-Planck-Landau equation in the literature. Less is known about the Landau equation with Coulomb interactions, the case when $\gamma = -3$. In fact, its well-posedness remained an open problem for a long time due to the analogy of the equation to a diffusion-reaction equation. Only recently Guillen and Silvestre proved in [12] that if the initial data is C^1 , bounded above by a Gaussian function and in particular with finite Fisher information, then there exists a unique strictly positive solution bounded above by a Gaussian. This solution enjoys conservations of mass, momentum and energy, as well as decay of entropy and Fisher information, although the decay rate is not known, unlike the Maxwellian molecules case.

Numerically approximating the Landau equation presents a difficult problem due to its high complexity. It is well-known that grid-based numerical methods such as finite difference/volume methods scale exponentially with dimension, which makes them not efficient for solving the full 6-dimensional Vlasov-Landau equation. This motivates developing dimension-agnostic numerical methods such as particle methods.

One such method for solving the Fokker-Planck equation was recently proposed by Shen et al. in [20] and Boffi and Vanden-Eijnden in [3], termed score-based transport modeling (SBTM) due to its similarity to score-based diffusion modeling. The authors took the gradient flow perspective on the Fokker-Planck equation and proposed a deterministic particle method for solving it using a neural network s to approximate the score $\nabla_v \log f$. In particular, they showed that as the score matching loss

$$\int_{\mathbb{R}^d} |s - \nabla_v \log f|^2 f \, dv$$

goes to zero, the Kullback-Leibler (KL) divergence between the approximate solution and the true solution goes to zero. This gives a theoretical guarantee for using the score matching method for solving the Fokker-Planck equation. Following this line of thinking, it is clear that the neural network is not necessarily the only choice to approximate the score. In fact, any rich enough family of functions can be used. In this vein, Maoutsa et al. [16] used reproducible kernel Hilbert spaces to minimize the score matching loss. The score matching idea has also been applied to other equations such as the McKean-Vlasov equation [19, 15].

Recently, score matching has been widely used in diffusion generative modeling for image generation [21, 22]. Chen et al. [8] showed that an L^2 approximation of the score guarantees fast convergence to the steady-state of the reverse diffusion process used for image generation. However, the idea of score matching is not new in scientific computing community – in 1990 Degond and Mustieles [10] developed the diffusion-velocity method for the heat equation, where the score was approximated using the traditional kernel-based particle method. Approximating the score with a neural network instead of kernels may be preferred due to the $O(n)$ vs $O(n^2)$ computational time where n is the total number of particles, and due to better accuracy in high dimensions. Unlike generative modeling, SBTM does not have the costly score pre-training step, with the exception of matching the score of the initial data, which takes seconds on a single CPU.

Another deterministic particle method that we explore in this work was proposed by Carrillo et al. in [6, 5], termed the blob method, first for the Fokker-Planck and porous medium equation and then for the Landau equation. The starting point is to recognize that $\nabla_v \log f$ is the gradient of the functional derivative of the Boltzmann entropy:

$$\nabla_v \log f = \nabla_v \frac{\delta H}{\delta f}, \quad H(f) = \int_{\mathbb{R}^d} f \log f \, dv.$$

If f is approximated by the empirical measure of n particles (Dirac measures), entropy is undefined, so regularized entropy is used instead:

$$H_\varepsilon(f) = \int_{\mathbb{R}^d} f \log(\phi_\varepsilon * f) \, dv,$$

for some symmetric and positive kernel ϕ_ε , often taken as the Gaussian. The authors showed that under the proposed numerical method the regularized entropy H_ε decays over time, mimicking the decay of the entropy H for the true solution.

The main contribution of this work is to show that the SBTM method developed for the Fokker-Planck equation can be used to approximate the Landau equation, with any value of γ . Furthermore, the method can preserve some important physical properties of the Landau equation, such as the conservation of mass, momentum, and energy, and the dissipation of estimated entropy. We also obtain a KL/relative entropy bound similar to that for the Fokker-Planck equation in [3, 19]. We perform comprehensive numerical experiments in different dimensions, with particular emphasis on comparing the proposed method with the blob method [5].

During the completion process of this manuscript, we became aware of the independent work [13]. The method proposed in [13] is very similar to ours, with a few key differences which we summarize below:

- The theoretical result obtained in [13] is based on the assumption that the Landau equation is posed on the torus \mathbb{T}^d rather than \mathbb{R}^d . This is a restrictive assumption under which many terms become bounded, for example, $|\nabla \log f(t, v)|$ is generally not a bounded function in \mathbb{R}^d , even when $f(t, v)$ is a Gaussian distribution. Unlike this work, our proof and method apply to the unbounded domain, which is the realistic setting for the Landau equation.
- Our numerical experiments focus on the comparison of the SBTM and the blob method in [5]. In particular, we demonstrate that for the simple isotropic BKW solution, both methods can perform well with the same number of particles. In fact, it is the anisotropic solution that distinguishes the power of SBTM over the blob method. Moreover, the SBTM method outperforms the blob method in higher dimensions, making it a potentially better choice for solving the full spatially inhomogeneous Vlasov-Landau equation. On the other hand, the work [13] mainly considers the test problems from [5] for which both methods perform well.

The outline of this paper is as follows. In Section 2 we review some theoretical properties of the Landau equation. These results are useful later to establish the theoretical estimate of the proposed numerical method. The score-based transport modeling (SBTM) method is introduced in Section 3. In Section 4 we prove a quantitative result justifying the use of the score matching loss. Section 5 contains the numerical experiments. We conclude in Section 6 with an outlook comparing SBTM to the blob method and describing future work.

2. Properties of the Landau equation

Our starting point is to write the Landau Equation (1) in the form of a continuity equation – a fundamental tool in the theory of optimal transport [25]. If a PDE can be written as a continuity equation, one can obtain its weak solution by solving a system of characteristic ODEs, where each ODE corresponds to a particle. The particle perspective is useful both for intuitive understanding and for numerical methods. See Appendix for a quick review on the continuity equation and its particle solution.

To be consistent with the optimal transport literature, in the rest of the paper we use the following notation:

- We use $u_t(x)$ to denote the density function $f(t, v)$, with x instead of v as the independent variable.
- We normalize u_t to be a probability density, so that $\int_{\mathbb{R}^d} u_t(x) dx = 1$. Additionally, we do not distinguish between a density and its distribution.

In this notation the Landau Equation (1) can be written as

$$\frac{\partial u_t}{\partial t} + \nabla \cdot (v_t u_t) = 0, \tag{2.1}$$

$$v_t(x) = - \int_{\mathbb{R}^d} A(x - y) [\nabla \log u_t(x) - \nabla \log u_t(y)] u_t(y) dy, \tag{2.2}$$

$$A_{i,j}(z) = |z|^\gamma (|z|^2 \delta_{i,j} - z_i z_j).$$

The following proposition combines Proposition 4 in [26], Theorem 2.3 in [12] and Theorem 1.2 in [12], where the authors established existence and uniqueness of solutions to the Landau equation, as well as no finite-time blowup.

THEOREM 2.1 (Existence, uniqueness, and regularity). *In case of the Maxwellian molecules ($\gamma = 0$ in (2)), if the initial data u_0 has finite energy*

$$\int_{\mathbb{R}^d} |x|^2 du_0(x) < \infty,$$

the Landau Equation (2) has a unique solution u_t defined for all $t \geq 0$. Moreover, for all $t > 0$, u_t is bounded and belongs to $C^\infty(\mathbb{R}^d)$. If u_0 is sub-Gaussian, then u_t is sub-Gaussian for all t :

$$u_0(x) \leq C_1 \exp(-\beta|x|^2) < \infty \implies u_t(x) \leq C_2 \exp(-\beta|x|^2) < \infty, \quad C_1, C_2 > 0.$$

In case of the Coulomb interactions, $\gamma = -3$ in (2) with $d = 3$, if u_0 is C^1 , bounded above by a Gaussian

$$u_0(x) \leq C_0 \exp(-\beta|x|^2),$$

and has finite Fisher information

$$\int_{\mathbb{R}^d} |\nabla \log u_0|^2 du_0(x) < \infty,$$

then there exists a unique smooth solution u_t that does not blow up in finite time.

The next proposition shows that the Landau equation with Maxwellian molecules can be written in the form of the Fokker-Planck equation with a time-and-space-dependent diffusion matrix. Note that this derivation already appeared in [26] but

with several rescaling and normalization assumptions. We redid the calculation here for convenience of presentation, as our later proof will rely on this proposition.

First of all, we recall that the solution to the Landau Equation (2) preserves mass, momentum and energy:

$$\int_{\mathbb{R}^d} u_t(x) dx \equiv 1, \quad V := \int_{\mathbb{R}^d} x u_t(x) dx = \int_{\mathbb{R}^d} x u_0(x) dx,$$

$$2E := \int_{\mathbb{R}^d} |x|^2 u_t(x) dx = \int_{\mathbb{R}^d} |x|^2 u_0(x) dx.$$

This property can be easily verified using a weak form of the Landau operator. Since our numerical solution also preserves this property, we defer our proof to the later section. We further define the (time-dependent) second order moment by matrix

$$\Sigma(t) := \int_{\mathbb{R}^d} x \otimes x u_t(x) dx.$$

With the notations V , E , and Σ , one can show the following:

PROPOSITION 2.1 (Fokker-Planck form). *The Landau Equation (2) with Maxwellian molecules ($\gamma = 0$) can be written in the Fokker-Planck form:*

$$\frac{\partial u_t}{\partial t} + \nabla \cdot (v_t u_t) = 0, \quad v_t(x) = b(x) - D_t(x) \nabla \log u_t,$$

$$b(x) = A * \nabla u_t = -(d-1)(x - V), \quad D_t(x) = A * u_t,$$

with the component of the matrix $D_t(x)$ given by

$$D_{i,j}(t, x) = \delta_{i,j} (|x|^2 + 2E - 2(V \cdot x)) + V_i x_j + V_j x_i - x_i x_j - \Sigma_{i,j}(t),$$

$$\Sigma_{i,j}(t) = \Sigma_{i,j}(\infty) - (\Sigma_{i,j}(\infty) - \Sigma_{i,j}(0)) e^{-4dt},$$

$$\Sigma_{i,j}(0) = \int_{\mathbb{R}^d} x_i x_j u_0(x) dx, \quad \Sigma_{i,j}(\infty) = \frac{1}{d} (\delta_{i,j} (2E - |V|^2) + dV_i V_j).$$

Proof. We must show that the velocity field v_t given in (2) can be written in the form $b - D \nabla \log u_t$. First, we split up v_t into two terms:

$$v_t = - \int_{\mathbb{R}^d} A(x - y) [\nabla \log u_t(x) - \nabla \log u_t(y)] u_t(y) dy$$

$$= (A * (\nabla u_t)) - (A * u_t) \nabla \log u_t,$$

where the first term is the drift term and the second term is the diffusion term. The drift term simplifies to

$$(A * (\nabla u_t))_i = \sum_j \partial_j A_{i,j} * u_t = -(d-1)(x_i - V_i).$$

As for the diffusion term, we first compute the convolution $A * u_t$:

$$(A * u_t)_{i,j} = \int_{\mathbb{R}^d} \delta_{i,j} |x - y|^2 u_t(y) dy - \int_{\mathbb{R}^d} (x_i - y_i)(x_j - y_j) u_t(y) dy$$

$$= \delta_{i,j} (|x|^2 + 2E - 2V \cdot x) + V_i x_j + V_j x_i - x_i x_j - \Sigma_{i,j}.$$

By a straightforward calculation, one can show that Σ satisfies

$$\Sigma_{i,j}(t) = \Sigma_{i,j}(\infty) - (\Sigma_{i,j}(\infty) - \Sigma_{i,j}(0))e^{-4dt}.$$

This completes the proof. □

We will also need the following elementary lemma, which is certainly known but, to our knowledge, has not explicitly appeared in the literature. A closely related estimate is given by Theorem 2.1 in [1].

LEMMA 2.1 (Boundedness of $A*u$). *For any distribution u not concentrated on a line, and Maxwellian collision kernel $A(z) = |z|^2 I_d - z \otimes z$, the matrix $A*u$ is uniformly bounded from below and bounded from above, i.e., there exists $\varepsilon > 0$ such that for all $x \in \mathbb{R}^d$,*

$$\varepsilon \leq \|A*u(x)\| \leq 2E + |x - V|^2, \quad 2E = \int_{\mathbb{R}^d} |x|^2 du(x), \quad V = \int_{\mathbb{R}^d} x du(x),$$

where $\|\cdot\|$ denotes the matrix 2-norm, and ε can be taken to be

$$\varepsilon = \text{tr}(\text{cov}(u)) - \|\text{cov}(u)\|.$$

In particular, the bounds above hold for any continuous density.

Proof. We denote $x' = x - V$ and compute

$$A*u(x) = I_d(|x'|^2 + 2E - |V|^2) - (x' \otimes x' + \Sigma - V \otimes V).$$

By triangle inequality, and noting that $\|x' \otimes x'\| = |x'|^2$, we have

$$\|A*u(x)\| \geq |x'|^2 + \text{tr}(\Sigma - V \otimes V) - |x'|^2 - \|\Sigma - V \otimes V\| = \text{tr}(\text{cov}(u)) - \|\text{cov}(u)\| > 0.$$

The last inequality is strict because $\text{tr}(\text{cov}(u))$ is the sum of the eigenvalues of $\text{cov}(u)$, whereas $\|\text{cov}(u)\|$ is the largest eigenvalue. The two quantities are equal if and only if u is concentrated on a line.

To get the upper bound we recall that Σ is positive semi-definite and compute

$$\|A*u(x)\| \leq |x'|^2 + 2E - |V|^2 - \inf_{|z|=1} \{z^T \Sigma z + |x \cdot z|^2\} + |V|^2 \leq 2E + |x - V|^2.$$

□

Lastly, we will need the following key estimate from a recent work [7].

THEOREM 2.2. *Let u_t solve the Landau equation with Maxwellian molecules, that is the case $\gamma = 0$ in any dimension, with initial data $u_0 \in W^{2,1} \cap W^{2,\infty}$. Assume further that the initial data u_0 satisfies the estimates*

$$u_0(x) \leq C_1 \exp(-|x|^2/C_1), \quad \text{and} \quad |\nabla \log u_0(x)| \leq C_2(1 + |x|).$$

Then for any $t \leq T$, the score function grows at most linearly in x , i.e., there exists a universal constant $C > 0$ such that

$$|\nabla \log u_t| \leq C(1 + |x| + \sqrt{t}).$$

This theorem is a major estimate in [7]. Combining with the estimates on $\nabla^2 \log u_t(x)$, the authors can establish entropic propagation of chaos from particle systems towards the Landau equation with Maxwellian molecules. We refer the readers to the proof in [7].

3. Particle methods for the Landau equation

As outlined in the Appendix, in order to (weakly) solve the continuity equation

$$\frac{\partial u_t}{\partial t} + \nabla \cdot (v_t u_t) = 0$$

with particles, the standard technique is to assume that the particle solution is given by

$$u_t(x) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(t)}(x),$$

and integrate the ODEs

$$\frac{dX_i}{dt} = v_t(X_i).$$

The challenge to solve the Landau Equation (2) lies in the approximation of the velocity field v_t . Since this velocity field mainly depends on the solution u_t through the score $\nabla \log u_t$, we aim to find an approximation $s_t(x)$ to the score $\nabla \log u_t(x)$. The approximate velocity v_t then becomes

$$v_t[s_t](x) = - \int_{\mathbb{R}^d} A(x-y)[s_t(x) - s_t(y)]u_t(y) dy. \tag{3.1}$$

Using particles $\{X_i\}_{i=1}^n$, this is

$$v_t[s_t](X_i) = - \frac{1}{n} \sum_{j=1}^n A(X_i - X_j)[s_t(X_i) - s_t(X_j)].$$

Before addressing the issue of how to estimate the score, we assume at this point that we have at our disposal a sufficiently regular time-dependent vector field $s_t(x)$. We establish the following result.

PROPOSITION 3.1. *Let $s_t(x)$ be any sufficiently regular time-dependent vector field. Consider the continuity equation*

$$\begin{aligned} \frac{\partial u_t}{\partial t} + \nabla \cdot (v_t u_t) &= 0, \\ v_t(x) &= - \int_{\mathbb{R}^d} A(x-y)[s_t(x) - s_t(y)]du_t(y), \end{aligned}$$

with initial density $u_0(x)$.

- (1) *The density solution $u_t(x)$ to the above equation conserves mass, momentum and energy. Furthermore, the estimated entropy decays. That is,*

$$\int_{\mathbb{R}^d} u_t dx = \int_{\mathbb{R}^d} u_0 dx, \quad \int_{\mathbb{R}^d} x u_t dx = \int_{\mathbb{R}^d} x u_0 dx, \quad \int_{\mathbb{R}^d} |x|^2 u_t dx = \int_{\mathbb{R}^d} |x|^2 u_0 dx,$$

$$\frac{d}{dt} \mathcal{E}_t \leq 0, \quad \mathcal{E}_t := \int_0^t \int_{\mathbb{R}^d} s_\tau(x) \cdot v_\tau(x) u_\tau(x) dx d\tau.$$

- (2) The particle solution $\frac{1}{n} \sum_{i=1}^n \delta_{X_i(t)}(x)$ with $\frac{dX_i}{dt} = v_t(X_i)$ to the above equation conserves mass, momentum, and energy. Furthermore, the estimated entropy decays. That is,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n X_i(t) &= \frac{1}{n} \sum_{i=1}^n X_i(0), & \frac{1}{n} \sum_{i=1}^n |X_i(t)|^2 &= \frac{1}{n} \sum_{i=1}^n |X_i(0)|^2. \\ \frac{d}{dt} \mathcal{E}_t &\leq 0, & \mathcal{E}_t &:= \frac{1}{n} \int_0^t \sum_{i=1}^n s_\tau(X_i) \cdot v_\tau(X_i) d\tau. \end{aligned}$$

- (3) The forward-Euler time-discretized particle solution $X_i^{k+1} = X_i^k + \Delta t v_t(X_i^k)$ conserves mass and momentum, and decays estimated entropy:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n X_i^{k+1} &= \frac{1}{n} \sum_{i=1}^n X_i^k, \\ \mathcal{E}^{k+1} &\leq \mathcal{E}^k, & \mathcal{E}^k &:= \frac{\Delta t}{n} \sum_{j=1}^k \sum_{i=1}^n s(X_i^j) \cdot v(X_i^j), \end{aligned}$$

where $\{X_i^k\}_{i=1}^n$ are the particle locations at time $t_k = k\Delta t$.

Proof. To prove the first two items it suffices to prove that any weak solution to the continuity equation above conserves mass, momentum and energy, and has a non-positive entropy decay rate. To that end, let $u_t(x)$ be a weak solution and let $\phi: \mathbb{R}^d \rightarrow \mathbb{R}$ be a test function. By the definition of a weak solution,

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}^d} \phi(x) du_t(x) &= \int_{\mathbb{R}^d} \nabla \phi(x) \cdot v_t(x) du_t(x) \\ &= - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \nabla \phi(x)^T A(x-y) [s(x) - s(y)] du_t(y) du_t(x) \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \nabla \phi(y)^T A(x-y) [s_t(x) - s_t(y)] du_t(y) du_t(x) \\ &= -\frac{1}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} [\nabla \phi(x) - \nabla \phi(y)]^T A(x-y) [s_t(x) - s_t(y)] du_t(y) du_t(x). \end{aligned}$$

Taking $\phi(x) = 1, x, |x|^2$, we get the desired conservations for u_t , using the fact that $A(z)z = 0$. Taking $\nabla \phi = s_t$ shows the non-positivity of the estimated entropy decay rate

$$\begin{aligned} \int_{\mathbb{R}^d} s_t(x) \cdot v_t(x) du_t(x) &= -\frac{1}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} [s_t(x) - s_t(y)]^T A(x-y) [s_t(x) - s_t(y)] du_t(y) du_t(x) \\ &\leq 0, \end{aligned}$$

since A is positive semi-definite.

To obtain the conservation of momentum for the time-discretized particle solution, we compute

$$\begin{aligned} \frac{1}{n} \sum_i X_i^{k+1} &= \frac{1}{n} \sum_i X_i^k - \frac{\Delta t}{n^2} \sum_{i,j} A(X_i^k - X_j^k) [s(X_i^k) - s(X_j^k)] \\ &= \frac{1}{n} \sum_i X_i^k + \frac{\Delta t}{n^2} \sum_{i,j} A(X_i^k - X_j^k) [s(X_i^k) - s(X_j^k)] \\ &= \frac{1}{n} \sum_i X_i^k. \end{aligned}$$

The second-to-last line is obtained by interchanging i and j and the symmetry of A . This establishes the conservation of momentum. The mass conservation is obvious because each particle carries the same weight $1/n$. To obtain the decay of estimated entropy we compute

$$\begin{aligned} \mathcal{E}^{k+1} - \mathcal{E}^k &= \frac{\Delta t}{n} \sum_{j=1}^{k+1} \sum_{i=1}^n [s(X_i^j) \cdot v(X_i^j)] - \frac{\Delta t}{n} \sum_{j=1}^k \sum_{i=1}^n [s(X_i^j) \cdot v(X_i^j)] \\ &= \frac{\Delta t}{n} \sum_{i=1}^n [s(X_i^{k+1}) \cdot v(X_i^{k+1})] \\ &= -\frac{\Delta t}{n^2} \sum_{i,j=1}^n s(X_i^{k+1})^T A(X_i^{k+1} - X_j^{k+1}) [s(X_i^{k+1}) - s(X_j^{k+1})] \\ &= -\frac{\Delta t}{2n^2} \sum_{i,j=1}^n [s(X_i^{k+1}) - s(X_j^{k+1})]^T A(X_i^{k+1} - X_j^{k+1}) [s(X_i^{k+1}) - s(X_j^{k+1})] \leq 0. \end{aligned}$$

□

REMARK 3.1. Note that above discussion is completely independent of the choice of $s_t(x)$. As long as it is a reasonable approximation to the score $\nabla \log u_t(x)$, one can obtain an approximate solution to the Landau equation, either in the continuous form or in the particle form, such that the conservation property and an estimated entropy decay property would hold. This viewpoint can put a large class of particle methods under the same umbrella, such as the blob method [5] or the SBTM method to be discussed below.

3.1. Score based transport modeling (SBTM). We now come to the approximation of the score for the Landau equation. Motivated by the score matching idea due to [14, 16] and subsequently used in [3], we propose to find the approximation $s(x)$ by minimizing the loss

$$\int_{\mathbb{R}^d} (\nabla \log u - s)^T (A * u) (\nabla \log u - s) u \, dx := \int_{\mathbb{R}^d} |\nabla \log u - s|_{A * u}^2 u \, dx,$$

where A is the kernel matrix given by (2).

In fact,

$$\begin{aligned} \int_{\mathbb{R}^d} |\nabla \log u - s|_{A * u}^2 u \, dx &= \int_{\mathbb{R}^d} (|\nabla \log u|_{A * u}^2 + |s|_{A * u}^2) u \, dx - 2 \int_{\mathbb{R}^d} s^T (A * u) (\nabla \log u) u \, dx \\ &= \int_{\mathbb{R}^d} (|\nabla \log u|_{A * u}^2 + |s|_{A * u}^2) u \, dx - 2 \int_{\mathbb{R}^d} s^T (A * u) \nabla u \, dx \\ &= \int_{\mathbb{R}^d} (|\nabla \log u|_{A * u}^2 + |s|_{A * u}^2) u \, dx + 2 \int_{\mathbb{R}^d} u \nabla \cdot ((A * u) s) \, dx, \end{aligned}$$

where we used integration by parts on the second-to-last line. Since the term $\int_{\mathbb{R}^d} |\nabla \log u|_{A * u}^2 u \, dx$ is a constant in s , minimizing the score matching loss

$$\int_{\mathbb{R}^d} |\nabla \log u - s|_{A * u}^2 u \, dx \tag{3.2}$$

is equivalent to minimizing

$$\int_{\mathbb{R}^d} (|s|_{A * u}^2 + 2 \nabla \cdot ((A * u) s)) u \, dx. \tag{3.3}$$

While $\nabla \log u$ and loss (3.1) are undefined for a sum of Dirac deltas $\frac{1}{n} \sum_{i=1}^n \delta_{X_i}(x)$, loss (3.1) is well-defined and is equal to

$$\frac{1}{n} \sum_{i=1}^n [s(X_i)^T D(X_i) s(X_i) + 2\nabla \cdot (D(X_i) s(X_i))], \quad D(X_i) := \frac{1}{n} \sum_{j=1}^n A(X_i - X_j).$$

Moreover, if X_1, \dots, X_n are i.i.d. samples from u , then by the law of large numbers

$$\frac{1}{n} \sum_{i=1}^n [s(X_i)^T D(X_i) s(X_i) + 2\nabla \cdot (D(X_i) s(X_i))] \rightarrow \mathbb{E}_u (|s|_{A * u}^2 + 2\nabla \cdot ((A * u) s)),$$

almost surely as $n \rightarrow \infty$.

Since replacing matrix D with the identity does not change the minimizer $s = \nabla \log u$, and empirically we observed no difference in accuracy between using the weighted and unweighted loss, to save the computational cost, we use the unweighted loss in later numerical experiments

$$\frac{1}{n} \sum_{i=1}^n [|s(X_i)|^2 + 2\nabla \cdot s(X_i)]. \tag{3.4}$$

Automatic differentiation can be used to compute the divergence of s but for efficiency reasons we use the *denoising trick* following [3]. The denoising trick is a way to approximate the divergence of a vector field, which makes the SBTM method approximately 100 times faster. As the authors of [3] point out, using the denoising trick prevents overfitting, so even leaving the computational time aside it is advisable to use the denoising trick. The denoising trick is the simple observation that given $Z \sim N(0, I_d)$ and $\alpha > 0$,

$$\lim_{\alpha \rightarrow 0} (2\alpha)^{-1} \mathbb{E}[s(x + \alpha Z) \cdot Z - s(x - \alpha Z) \cdot Z] = \nabla \cdot s(x).$$

So the divergence of *any vector field* s can be approximated by a finite sum

$$\frac{1}{p} \sum_{j=1}^p (2\alpha)^{-1} [s(x + \alpha Z_j) \cdot Z_j - s(x - \alpha Z_j) \cdot Z_j],$$

where Z_j are iid draws from the standard Gaussian distribution. In practice we set $p=1$. Thus, the loss we minimize is

$$L(s) = \frac{1}{n} \sum_{i=1}^n [|s(X_i)|^2 + \alpha^{-1} (s(X_i + \alpha Z) - s(X_i - \alpha Z)) \cdot Z]. \tag{3.5}$$

To approximate $\nabla \log u$ well using the minimization problem

$$s = \operatorname{argmin}_{s \in \mathcal{F}} L(s),$$

the class of functions \mathcal{F} must be rich enough to contain $\nabla \log u$. The authors of [3] propose to take \mathcal{F} to be the family of neural networks since they are universal approximators, and to minimize the loss $L(s)$ using gradient descent.

The pseudocode for solving the Landau equation is as follows.

SBTM($u_0, n, t_0, t_{\text{end}}, \Delta t, K$)

- 1 $t := t_0$
- 2 sample $\{X_i\}_{i=1}^n$ from u_0
- 3 initialize NN: $s \approx \arg \min_{s \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n |s(X_i) - \nabla \log u_0(X_i)|^2$
- 4 while $t < t_{\text{end}}$
- 5 + $t := t + \Delta t$
- 6 Optimize s : do K gradient descent steps on $L(s)$ in (3.1)
- 7 for $i = 1, \dots, n$
- 8 + $X_i := X_i - \frac{\Delta t}{n} \sum_{j=1}^n A(X_i - X_j)[s(X_i) - s(X_j)]$
- 9 output particle locations X_1, \dots, X_n

The initialization is done by performing gradient descent steps until the loss is below some pre-specified threshold.

4. Kullback–Leibler divergence bound

How good is the score matching approach? In this section, we try to answer this question by quantifying the Kullback–Leibler (KL) divergence between the exact solution of the Landau equation and that of SBTM, for the semi-discretized scheme. The KL divergence or the relative entropy can quantify the distance between two probability densities. It is defined as

$$\text{KL}(u_1||u_2) = \begin{cases} \int_{\mathbb{R}^d} \log \frac{u_1}{u_2} \, du_1(x), & \text{if } u_1 \text{ is abolutely continuous w.r.t. } u_2, \\ +\infty, & \text{otherwise.} \end{cases}$$

By convexity, one can easily see that $\text{KL}(u_1||u_2) \geq 0$ and $\text{KL}(u_1||u_2) = 0$ if and only if $u_1 = u_2$.

Before analyzing the Landau equation, we review a similar bound from the work of Boffi and Vanden-Eijnden [3].

4.1. KL bound for the Fokker-Planck equation.

THEOREM 4.1. *Fix $b_t(x), D_t(x)$, such that D_t is a positive semi-definite matrix. Let $s_t(x) \in \mathbb{R}^d$ be sufficiently regular and u_t and u_t^* satisfy the continuity equations*

$$\begin{aligned} \frac{\partial u_t^*}{\partial t} + \nabla \cdot (v_t^* u_t^*) &= 0, & v_t^* &= b_t - D_t \nabla \log u_t^*, \\ \frac{\partial u_t}{\partial t} + \nabla \cdot (v_t u_t) &= 0, & v_t &= b_t - D_t s_t. \end{aligned}$$

Then the KL-divergence is controlled by the score matching loss

$$\frac{d}{dt} \text{KL}(u_t||u_t^*) \leq \frac{1}{2} \int_{\mathbb{R}^d} |s_t - \nabla \log u_t|_{D_t}^2 u_t \, dx.$$

Proof. This follows immediately from the time evolution of the KL/relative entropy between u_t and u_t^* . See for instance the proof of Proposition 1 in [3]. □

By combining the above theorem with Proposition 2.1, we immediately obtain the following corollary.

COROLLARY 4.1. *Let u_t^* satisfy the Landau equation with Maxwellian molecules*

$$\frac{\partial u_t^*}{\partial t} + \nabla \cdot (v_t^* u_t^*) = 0, \quad v_t^*(x) = - \int_{\mathbb{R}^d} A(x-y)[\nabla \log u_t^*(x) - \nabla \log u_t^*(y)] u_t^*(y) \, dy,$$

and let u_t satisfy the continuity equation

$$\frac{\partial u_t}{\partial t} + \nabla \cdot (v_t u_t) = 0, \quad v_t(x) = b(x) - D_t(x) s_t(x),$$

where $b(x)$ and $D_t(x)$ are as in Proposition 2.1. Further, assume that the initial first and second moments match, i.e.,

$$\int_{\mathbb{R}^d} x u_0(x) dx = \int_{\mathbb{R}^d} x u_0^*(x) dx, \quad \int_{\mathbb{R}^d} x_i x_j u_0(x) dx = \int_{\mathbb{R}^d} x_i x_j u_0^*(x) dx.$$

Then the time-derivative of the KL-divergence is controlled by the score matching loss

$$\frac{d}{dt} \text{KL}(u_t || u_t^*) \leq \frac{1}{2} \int_{\mathbb{R}^d} |s_t - \nabla \log u_t|_{D_t}^2 u_t dx.$$

Proof. Because of conservations, the diffusion matrix $D_t(x)$ and the drift vector $b(x)$ match for all x, t . We also need to show that $D_t(x)$ is positive semi-definite. This follows by noting that

$$D_t(x) = A * u_t^*(x),$$

and recalling that A is positive semi-definite. □

This can be turned into a numerical method for solving the Landau equation with Maxwellian molecules by using $v_t = b_t - D_t s_t$ as the velocity field in the particle solution. However, this method does not preserve momentum and energy. Furthermore, this numerical scheme heavily relies on rewriting the Landau equation in the Fokker-Planck form, which is not possible for general kernels A , and specifically is not possible for the Coulomb kernel. Because of these reasons, we prove a similar inequality to Theorem 4.1 for the particle solution using the velocity field (3).

4.2. KL bound for the Landau equation. Let u_t^* be a C^1 density satisfying the Landau Equation (2) with Maxwellian molecules ($\gamma = 0$)

$$\begin{aligned} \frac{\partial u_t^*}{\partial t} + \nabla \cdot (v_t^* u_t^*) &= 0, \\ v_t^*(x) &= - \int_{\mathbb{R}^d} A(x-y) (\nabla \log u_t^*(x) - \nabla \log u_t^*(y)) u_t^*(y) dy, \end{aligned}$$

and let $s_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a time-dependent vector field and u_t be a C^1 density satisfying the approximate Landau equation

$$\begin{aligned} \frac{\partial u_t}{\partial t} + \nabla \cdot (v_t u_t) &= 0, \\ v_t(x) &= - \int_{\mathbb{R}^d} A(x-y) (s_t(x) - s_t(y)) u_t(y) dy. \end{aligned}$$

For both equations, we assume that they are subject to the same initial condition $u_0 = u_0^*$, so $\text{KL}(u_0 || u_0^*) = 0$.

Since the momentum and energy are conserved in both equations (by properties of the Landau equation and Proposition 3.1), we denote them by V and E respectively:

$$V = \int_{\mathbb{R}^d} x u_t(x) dx = \int_{\mathbb{R}^d} x u_t^*(x) dx, \quad 2E = \int_{\mathbb{R}^d} |x|^2 u_t(x) dx = \int_{\mathbb{R}^d} |x|^2 u_t^*(x) dx.$$

Additionally, we denote the (time-dependent) second moment by matrices Σ and Σ^* :

$$\Sigma(t) = \int_{\mathbb{R}^d} x \otimes x u_t(x) dx, \quad \Sigma^*(t) = \int_{\mathbb{R}^d} x \otimes x u_t^*(x) dx.$$

We first establish an intermediate result about the KL divergence.

PROPOSITION 4.1. *The KL divergence between u_t and u_t^* is controlled by the score matching loss plus an extra term that depends on the difference between second moments of u_t and u_t^* :*

$$\begin{aligned} \frac{d}{dt} KL(u_t || u_t^*) \leq & 2 \int_{\mathbb{R}^d} |\nabla \log u_t - s_t|_{A^* u_t}^2 u_t dx \\ & + \int_{\mathbb{R}^d} (\nabla \log u_t^*)^T (\Sigma - \Sigma^*) (\nabla \log u_t - \nabla \log u_t^*) u_t dx. \end{aligned}$$

Proof. We omit the subscript t throughout this proof to simplify notation. We define an auxillary vector field

$$\bar{v}(x) := - \int_{\mathbb{R}^d} A(x-y) (\nabla \log u(x) - \nabla \log u(y)) u(y) dy,$$

and compute

$$\begin{aligned} \frac{d}{dt} KL(u || u^*) &= - \int_{\mathbb{R}^d} \frac{u}{u^*} \partial_t u^* dx + \int_{\mathbb{R}^d} \log \frac{u}{u^*} \partial_t u dx \\ &= \int_{\mathbb{R}^d} \frac{u}{u^*} \nabla \cdot (v^* u^*) dx - \int_{\mathbb{R}^d} \nabla \cdot (vu) \log \left(\frac{u}{u^*} \right) dx \\ &= - \int_{\mathbb{R}^d} v^* \cdot \nabla \left(\frac{u}{u^*} \right) u^* dx + \int_{\mathbb{R}^d} v \cdot \nabla \log \left(\frac{u}{u^*} \right) u dx \\ &= \int_{\mathbb{R}^d} (v - v^*) \cdot (\nabla \log u - \nabla \log u^*) u dx \\ &= \int_{\mathbb{R}^d} (v - \bar{v}) \cdot (\nabla \log u - \nabla \log u^*) u dx + \int_{\mathbb{R}^d} (\bar{v} - v^*) \cdot (\nabla \log u - \nabla \log u^*) u dx. \end{aligned} \tag{4.1}$$

We bound the first term in (4.2) with Young's inequality.

$$\begin{aligned} & \int_{\mathbb{R}^d} (v - \bar{v}) \cdot (\nabla \log u - \nabla \log u^*) u dx \\ &= \int_{\mathbb{R}^d} (\nabla \log u - s)^T (A^* u) \cdot (\nabla \log u - \nabla \log u^*) u dx \\ & \quad - \int_{\mathbb{R}^d} A^* (u (\nabla \log u - s)) \cdot (\nabla \log u - \nabla \log u^*) u dx \\ & \leq \int_{\mathbb{R}^d} |\nabla \log u - s|_{A^* u}^2 u dx + \frac{1}{4} \int_{\mathbb{R}^d} |\nabla \log u - \nabla \log u^*|_{A^* u}^2 u dx \\ & \quad - \int_{\mathbb{R}^d} A^* (u (\nabla \log u - s)) \cdot (\nabla \log u - \nabla \log u^*) u dx, \end{aligned}$$

where

$$\begin{aligned} & - \int_{\mathbb{R}^d} A^* (u (\nabla \log u - s)) \cdot (\nabla \log u - \nabla \log u^*) u dx \\ & \leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\nabla \log u(y) - s(y)|_{A(x-y)}^2 u(x) u(y) dx dy \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{4} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\nabla \log u(x) - \nabla \log u^*(x)|_{A(x-y)}^2 u(x)u(y) \, dx \, dy \\
 & = \int_{\mathbb{R}^d} |\nabla \log u - s|_{A^*u}^2 u \, dx + \frac{1}{4} \int_{\mathbb{R}^d} |\nabla \log u - \nabla \log u^*|_{A^*u}^2 u \, dx.
 \end{aligned}$$

Together we have

$$\int_{\mathbb{R}^d} (v - \bar{v}) \cdot \left(\nabla \log \frac{u}{u^*} \right) u \, dx \leq 2 \int_{\mathbb{R}^d} |\nabla \log u - s|_{A^*u}^2 u \, dx + \frac{1}{2} \int_{\mathbb{R}^d} \left| \nabla \log \frac{u}{u^*} \right|_{A^*u}^2 u \, dx. \tag{4.2}$$

Now we analyze the second term in (4.2). Using the fact that A is the Maxwellian kernel, we can write

$$\begin{aligned}
 v^* & = -(A * u^*) \nabla \log u^* - (d-1)(x - V), \\
 \bar{v} & = -(A * u) \nabla \log u - (d-1)(x - V),
 \end{aligned}$$

so the drift term in $\bar{v} - v^*$ cancels out:

$$\bar{v} - v^* = (A * u^*) \nabla \log u^* - (A * u) \nabla \log u.$$

Thus

$$\begin{aligned}
 & \int_{\mathbb{R}^d} (\bar{v} - v^*) \cdot (\nabla \log u - \nabla \log u^*) u \, dx \\
 & = \int_{\mathbb{R}^d} ((A * u^*) \nabla \log u^* - (A * u) \nabla \log u) \cdot (\nabla \log u - \nabla \log u^*) u \, dx \\
 & = - \int_{\mathbb{R}^d} |\nabla \log u - \nabla \log u^*|_{A^*u}^2 u \, dx + \int_{\mathbb{R}^d} (\nabla \log u^*)^T (A * (u^* - u)) (\nabla \log u - \nabla \log u^*) u \, dx. \tag{4.3}
 \end{aligned}$$

Combining (4.2) and (4.3), we have

$$\begin{aligned}
 \frac{d}{dt} KL(u||u^*) & \leq 2 \int_{\mathbb{R}^d} |\nabla \log u - s|_{A^*u}^2 u \, dx - \frac{1}{2} \int_{\mathbb{R}^d} |\nabla \log u - \nabla \log u^*|_{A^*u}^2 u \, dx \\
 & \quad + \int_{\mathbb{R}^d} (\nabla \log u^*)^T (A * (u^* - u)) (\nabla \log u - \nabla \log u^*) u \, dx,
 \end{aligned}$$

where the second term is non-positive because $A * u$ is positive semi-definite (in fact, strictly positive definite by Lemma 2.1).

Finally, we note that $A * (u^* - u) = \Sigma - \Sigma^*$. This can be seen directly by computing

$$\begin{aligned}
 (A * u)_{i,j}(x) & = \delta_{i,j}(|x|^2 + 2E - 2V \cdot x) + (x_i V_j + x_j V_i - x_i x_j) - \Sigma_{i,j}(t), \\
 (A * u^*)_{i,j}(x) & = \delta_{i,j}(|x|^2 + 2E - 2V \cdot x) + (x_i V_j + x_j V_i - x_i x_j) - \Sigma_{i,j}^*(t),
 \end{aligned}$$

and subtracting the two equations. □

The second term in Proposition 4.1 is hard to control a priori, but if the initial data is isotropic, it vanishes.

COROLLARY 4.2. *Suppose the initial condition u_0 is radially symmetric and s_t is in the form $s_t(x) = x \hat{s}_t(|x|)$ for some $\hat{s} : \mathbb{R} \rightarrow \mathbb{R}^d$. Then the KL divergence is controlled only by the score matching loss.*

$$\frac{d}{dt} KL(u_t||u_t^*) \leq 2 \int_{\mathbb{R}^d} |\nabla \log u_t - s_t|_{A^*u_t}^2 u_t \, dx.$$

Proof. (Proof sketch). It suffices to prove that u_t and u_t^* stay isotropic. Indeed, in that case $\Sigma(t)$ and $\Sigma^*(t)$ are scalar matrices and thus are equal because their trace is equal to $2E$, which is conserved. So the difference $\Sigma(t) - \Sigma^*(t)$ in Proposition 4.1 is zero.

So it suffices to show that the velocity fields v_t and v_t^* are radial if u_t and u_t^* are isotropic. It is a straightforward computation, using the assumed form of s_t . \square

The assumed form of s_t is not restrictive, because we can simply take $\hat{s}_t : \mathbb{R} \rightarrow \mathbb{R}^d$ to be the neural network.

If the initial data is not isotropic, we only get a short time horizon estimate, but $KL(u_t || u_t^*)$ still goes to 0 as the time-integrated score matching loss goes to zero.

THEOREM 4.2. *Fix $T > 0$ and suppose that the following conditions hold:*

- (1) *The initial data is in a Sobolev space*

$$u_0 \in W^{2,1} \cap W^{2,\infty},$$

and there exists a constant C_0 such that

$$|\nabla \log u_0(x)| < C_0(1 + |x|).$$

- (2) *There exist constants C'_0 and $\beta > 0$ such that the initial condition is sub-Gaussian*

$$u_0^*(x) = u_0(x) \leq C'_0 e^{-3\beta|x|^2}.$$

Then for $t \leq T$ small enough such that $KL(u_t || u_t^) \leq 4$, the following estimate holds*

$$KL(u_t || u_t^*) \leq 2e^{Mt} \int_0^t \int_{\mathbb{R}^d} |\nabla \log u_\tau - s_\tau|_{A^*u_\tau}^2 u_\tau \, dx \, d\tau,$$

where

$$M = d^2(1+T) \frac{16C_1}{\beta^2\varepsilon} (1+E) \left(3 + 2 \log \left(C_2 \sqrt{\frac{\pi}{\beta}} \right) \right)^2,$$

$$\varepsilon = \inf_{t \leq T} \{ \text{tr}(\text{cov}(u_t)) - \|\text{cov}(u_t)\| \} > 0,$$

with C_1, C_2 being some positive constants. In particular, as the time-integrated score matching loss goes to zero, the KL-divergence goes to 0:

$$KL(u_t || u_t^*) \rightarrow 0, \quad \text{as} \quad \int_0^t \int_{\mathbb{R}^d} |\nabla \log u_\tau - s_\tau|_{A^*u_\tau}^2 u_\tau \, dx \, d\tau \rightarrow 0.$$

Proof. We omit the subscript t for convenience of notation. Following the proof of Proposition 4.1, there remains to control the term

$$-\frac{1}{2} \int_{\mathbb{R}^d} |\nabla \log u - \nabla \log u^*|_{A^*u}^2 u \, dx + \int_{\mathbb{R}^d} (\nabla \log u^*)^T (\Sigma - \Sigma^*) (\nabla \log u - \nabla \log u^*) u \, dx.$$

By Lemma 2.1,

$$A^*u - \varepsilon I \geq 0.$$

Using Young’s inequality, we get

$$\begin{aligned}
 & -\frac{1}{2} \int_{\mathbb{R}^d} \left| \nabla \log \frac{u}{u^*} \right|_{A^*u}^2 u \, dx + \int_{\mathbb{R}^d} (\nabla \log u^*)^T (\Sigma - \Sigma^*) \left(\nabla \log \frac{u}{u^*} \right) u \, dx \\
 \leq & -\frac{1}{2} \int_{\mathbb{R}^d} \left| \nabla \log \frac{u}{u^*} \right|_{A^*u}^2 u \, dx + \frac{\varepsilon}{2} \int_{\mathbb{R}^d} \left| \nabla \log \frac{u}{u^*} \right|^2 u \, dx + \frac{1}{2\varepsilon} \int_{\mathbb{R}^d} |(\Sigma - \Sigma^*) \nabla \log u^*|^2 u \, dx \\
 \leq & \frac{1}{2\varepsilon} \|\Sigma - \Sigma^*\|^2 \int_{\mathbb{R}^d} |\nabla \log u^*|^2 u \, dx \\
 \leq & \frac{4C_1}{\varepsilon} (1+t)(1+E) \|\Sigma - \Sigma^*\|^2.
 \end{aligned} \tag{4.4}$$

The last inequality (and constant $C_1 > 0$) follows from Theorem 2.2 and conservation of energy.

We rewrite the difference of second moments in the form of a weighted L^1 norm.

$$\begin{aligned}
 \|\Sigma - \Sigma^*\|^2 & \leq \sum_{i,j=1}^d (\Sigma_{i,j} - \Sigma^*_{i,j})^2 = \sum_{i,j=1}^d \left(\int_{\mathbb{R}^d} x_i x_j (u - u^*) \, dx \right)^2 \\
 & \leq d^2 \left(\int_{\mathbb{R}^d} |x|^2 |u - u^*| \, dx \right)^2.
 \end{aligned} \tag{4.5}$$

We are now in a position to use the weighted CKP inequality from [4]

$$\int_{\mathbb{R}^d} \phi(x) |u(x) - u^*(x)| \, dx \leq \left(3 + 2 \log \int_{\mathbb{R}^d} e^{2\phi(x)} u^*(x) \, dx \right) \left(\text{KL}(u||u^*)^{1/2} + \frac{1}{2} \text{KL}(u||u^*) \right),$$

with $\phi(x) = \beta|x|^2$. By Theorem 2.1, there exists a constant C_2 such that

$$u^*(x) \leq C_2 e^{-3\beta|x|^2}.$$

Thus,

$$\int_{\mathbb{R}^d} e^{2\beta|x|^2} u^*(x) \, dx \leq C_2 \int_{\mathbb{R}^d} e^{-\beta|x|^2} \, dx = C_2 \sqrt{\frac{\pi}{\beta}}.$$

We have

$$\begin{aligned}
 \beta \int_{\mathbb{R}^d} |x|^2 |u - u^*| \, dx & \leq \left(3 + 2 \log \int_{\mathbb{R}^d} e^{2\beta|x|^2} u^*(x) \, dx \right) \left(\text{KL}(u||u^*)^{1/2} + \frac{1}{2} \text{KL}(u||u^*) \right) \\
 & \leq \beta C_3 \text{KL}(u||u^*)^{1/2}, \quad \text{for all } t \text{ such that } \text{KL}(u||u^*) \leq 4, \\
 C_3 & = \frac{2}{\beta} \left(3 + 2 \log \left(C_2 \sqrt{\frac{\pi}{\beta}} \right) \right).
 \end{aligned} \tag{4.6}$$

Combining estimates (4.2), (4.2) and (4.2), we get the differential inequality that holds for $t \leq T$ small enough that $\text{KL}(u||u^*) \leq 4$

$$\begin{aligned}
 \frac{d}{dt} \text{KL}(u||u^*) & \leq L(t) + \frac{4C_1}{\varepsilon} (1+E) d^2 C_3^2 (1+T) \text{KL}(u||u^*) \\
 & = L(t) + M \text{KL}(u||u^*),
 \end{aligned}$$

with

$$L(t) = 2 \int_{\mathbb{R}^d} |s - \nabla \log u|_{A^*u}^2 u \, dx, \quad M = \frac{4C_1}{\varepsilon} (1+E) d^2 C_3^2 (1+T).$$

By generalized Grönwall's inequality we obtain

$$\text{KL}(u||u^*) \leq e^{Mt} \int_0^t L(\tau) d\tau.$$

□

REMARK 4.1. KL divergence upper bounds several common distances, with total variation being the most notable, so Theorem 4.2 can be extended to other distances. See [11] for an overview of distances between probability densities.

5. Numerical experiments

We compare the proposed method with the traditional blob method [5] on several examples with Maxwellian and Coulomb collision kernels. All simulations and plots were done in Julia programming language with the code available at [GradientFlows.jl](https://github.com/GradientFlows/jl). The code is well-tested with both unit tests and integration tests running on GitHub Actions.

All numerical experiments were performed on the Doppio cluster of the Applied Mathematics Department at the University of Washington – a Silicon Mechanics Rack-form R2504.V6 with two 20-core Intel Xeon E5-2698 v4 processors at 2.20GHz, 512 GB of RAM, and two TITAN X (Pascal) GPUs with 12 GB of RAM each.

5.1. Hyperparameter choice. In the blob method, as the kernel bandwidth ε goes to either zero or infinity, the velocity of the particles in the blob method goes to zero, which indicates that there is a “sweet spot” value of ε . The authors of [5] used the mesh size to set the value of ε . Since we do random initialization instead of grid initialization, we follow the statistical literature [9] on kernel density estimation and choose ε to minimize MISE from the gradient of the true solution. We use Silverman's plug-in method:

$$\varepsilon = n^{-1/(d+6)} (\sigma_1 \cdot \dots \cdot \sigma_d)^{1/d},$$

where σ_i are the eigenvalues of the covariance matrix Σ of the sample. We suspect that while this choice is asymptotically optimal for estimating ∇u , it might not be optimal for estimating $\nabla \log u = \frac{\nabla u}{u}$.

In the SBTM, there are more hyperparameters, the most important being

- (1) the NN architecture: number and sizes of layers
- (2) learning rate η
- (3) optimizer
- (4) activation function
- (5) number K of GD steps at each time step
- (6) denoising constant α

We tuned all of the above with trial and error and chose $\eta = 4 \cdot 10^{-4}$, the Adam optimizer, the `softsign` activation function, $K = 25$, and $\alpha = 0.4$. The choice of α presents a bias-variance trade-off: higher values of α introduce bias but reduce the variance of the approximation of (3.1). We use a neural network with two hidden layers, each with 100 neurons, for the BKW example and one hidden layer with 100 neurons for other examples. We use a larger neural network for the BKW example because initial score is nearly singular at 0.

We found that different learning rates work best on different examples, with the most common issue being underfitting of the score in the first few iterations. Choosing

the wrong learning rate may lead to underfitting (top right subfigure of Figure 5.1), and it is hard to select the correct learning rate without having access to the true solution, because the true score matching loss can only be plotted if the true solution is known. Instead, an adaptive strategy may be used: at step 6 of the SBTM algorithm, instead of doing K GD steps, perform GD steps on the denoising loss $L(s)$ in (3.1) until the implicit loss (3.1) stops decaying.

The adaptive training strategy sidesteps performing gradient descent on the true implicit loss, which is slow, but uses the true implicit score to detect convergence of the neural network. Figure 5.1 demonstrates how a poor choice of the learning rate η and the number of GD steps K can lead to severe underfitting. The black line is the true score matching loss (3.1), the orange line is the true implicit loss (3.1), and the blue line is the denoising loss (3.1). Each time the particles move, the loss increases, and each step of GD the loss decreases. The adaptive strategy proposed above does well for a wide range of learning rates (compare the two bottom plots), while the non-adaptive strategy of taking K GD steps is very sensitive to the learning rate. When the solution gets closer to the steady state, fewer GD steps are needed to keep a good approximation of the score, which may lead to a significant speed-up if the adaptive strategy is used – compare the total number of epochs on the horizontal axis of the two subfigures on the left. While the adaptive strategy is promising, we defer its exploration to future work.

For density estimation we use the kernel bandwidth given by the Scott’s rule of thumb [18]:

$$H^* = n^{-1/(d+4)}\Sigma,$$

where Σ is the covariance of the sample. It is asymptotically optimal for minimizing MISE if the underlying distribution is Gaussian. See [9] for a comprehensive review of multivariate kernel density estimation.

5.2. Landau equation with Maxwellian kernel.

EXAMPLE 5.1 (BKW solution in dimension $d=3$). *The BKW solution is an isotropic analytic solution to the Landau equation with Maxwellian kernel. In general dimension d , it is given by*

$$A_{i,j}(z) = B(|z|^2\delta_{i,j} - z_i z_j), \quad u_t^*(x) = (2\pi K)^{-d/2} \exp\left(-\frac{|x|^2}{2K}\right) (P + Q|x|^2),$$

$$P = \frac{(d+2)K - d}{2K}, \quad Q = \frac{1 - K}{2K^2}, \quad K = 1 - \exp(-2B(d-1)t), \quad B = 1/24.$$

It is derived in Appendix A of [5]. We use $t_0=5.5$ to be slightly larger than

$$t_0^* = \frac{\log(d/2+1)}{2B(d-1)} = 5.497 \quad \text{for } d=3, B=1/24,$$

so that $P(t_0)$ is slightly positive. At $t=t_0^$ the score is singular. We take $\Delta t=0.01$ and $t_{end}=9.5$.*

REMARK 5.1. Since the BKW solution stays isotropic, second moments are automatically matched, up to the sampling error, on the order of $\frac{1}{\sqrt{n}}$. This is because the trace of the second moment matrix Σ is $2E$, and the energy \bar{E} is conserved. Thus, we do not expect to see a big discrepancy between blob and SBTM in terms of matching the second moments. The simplicity of Corollary 4.2 is another indication of the degeneracy of isotropic initial data for the Landau equation.

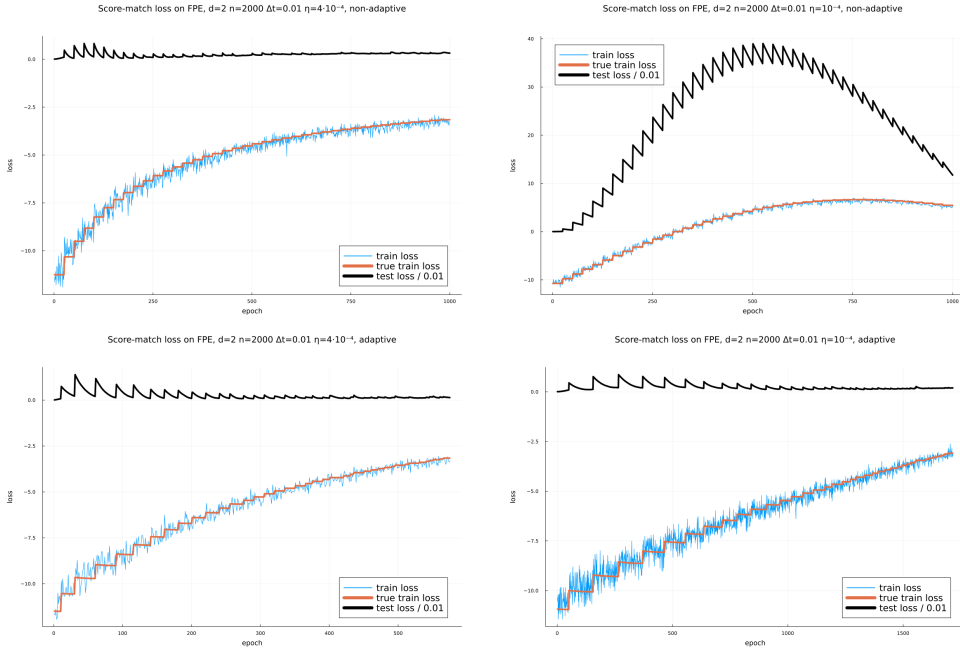


FIG. 5.1. Effect of underfitting, optimal $\eta = 4 \cdot 10^{-4}$ (left) vs low $\eta = 10^{-4}$ (right) and non-adaptive training strategy (top) vs adaptive training strategy (bottom).

Figure 5.2 shows several metrics that quantify how accurate the numerical solution is. The first plot shows the time evolution of the (1,1) entry of the covariance matrix Σ of the numerical solution and of the true solution. The second plot shows the Frobenius norm of the difference between the true second moment matrix $\Sigma^*(t_{\text{end}})$ and the second moment matrix $\Sigma(t_{\text{end}})$ of the numerical solution

$$\sum_{i,j=1}^d |\Sigma_{i,j} - \Sigma_{i,j}^*|^2, \quad \Sigma_{i,j} = \frac{1}{n} \sum_k (X_k)_i (X_k)_j,$$

for particles $\{X_k\}_{k=1}^n$. The third metric is the normalized true score matching error

$$\frac{\sum_i |\nabla \log u_t^*(X_i) - s(X_i)|^2}{\sum_i |\nabla \log u_t^*(X_i)|^2}.$$

The last metric is the L2 error of the reconstructed density against the true density

$$\left(\int_{\mathbb{R}^d} |u_t^*(x) - \frac{1}{n} \sum_j \phi_h(x - X_j)|^2 dx \right)^{1/2}.$$

The L2 error relies on the reconstructed density, as opposed to just the particle solution, so this plot may look very different depending on the choice of bandwidth. We also plot the conserved moments. The fifth plot in Figure 5.2 shows the norm of the difference between the sample mean at time t_0 and t_{end}

$$\left| \frac{1}{n} \sum_i X_i(t_{\text{end}}) - \frac{1}{n} \sum_i X_i(t_0) \right|.$$

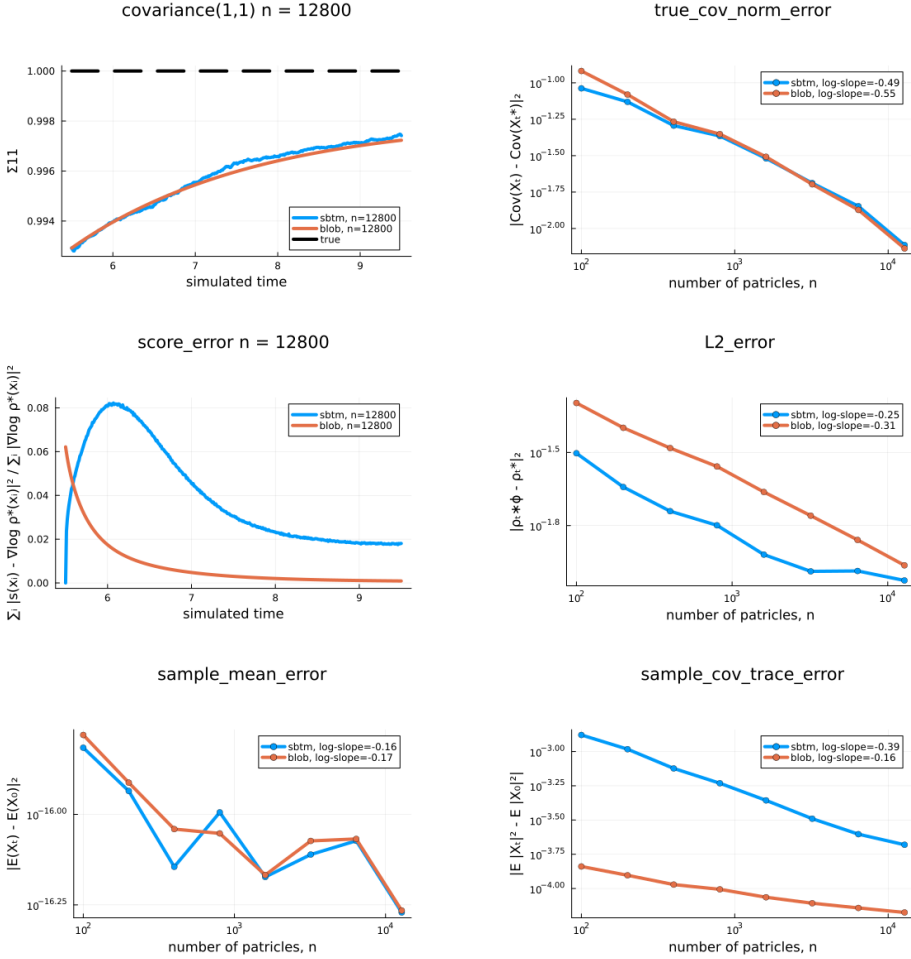


FIG. 5.2. (An isotropic) BKW solution to the Landau equation with Maxwellian kernel in dimension $d=3$. Both the blob method and SBTM do a decent job of approximating the solution.

As predicted by Proposition 3.1, this quantity is identically zero, up to floating point error. The sixth plot shows the difference between the sample kinetic energy at time t_0 and t_{end}

$$\left| \frac{1}{n} \sum_i |X_i(t_{\text{end}})|^2 - \frac{1}{n} \sum_i |X_i(t_0)|^2 \right|.$$

This quantity is not identically zero because of time discretization. But it is on the order of $\Delta t = 0.01$.

As described above, the score matching error is very sensitive to the learning rate. It is likely that the learning rate of $\eta = 4 \cdot 10^{-4}$ is too low for this example but we refrained from optimizing the hyperparameters for each individual example for the sake of a fair comparison.

Per Remark 5.1, in order to meaningfully assess the quality of the proposed numerical method in terms of matching the second moments, we must use non-isotropic

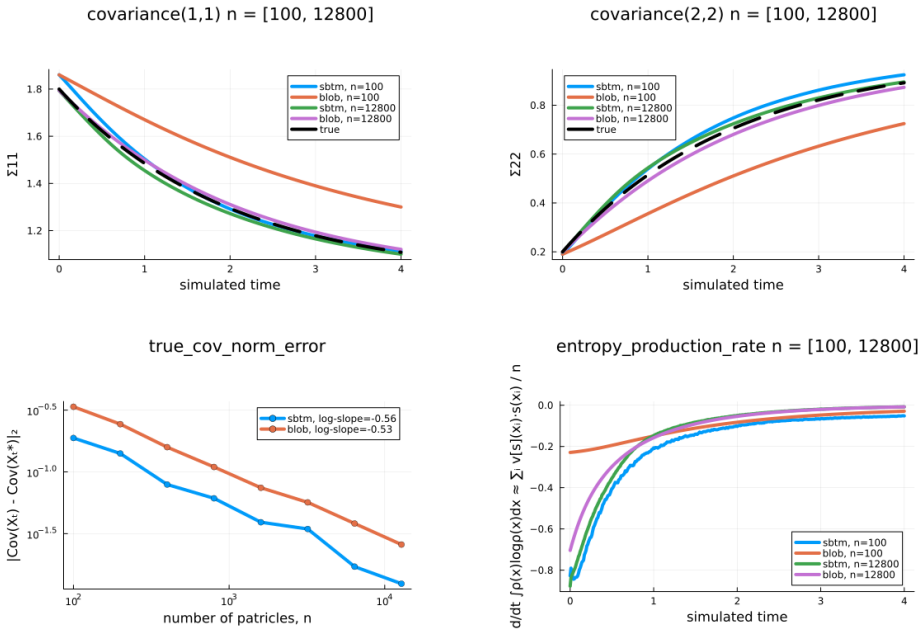


FIG. 5.3. (An anisotropic) solution to the Landau equation with Maxwellian kernel in dimension $d=3$. SBTM can match the covariance well even with $n=100$ particles, while the blob method needs a lot more particles to achieve the same accuracy.

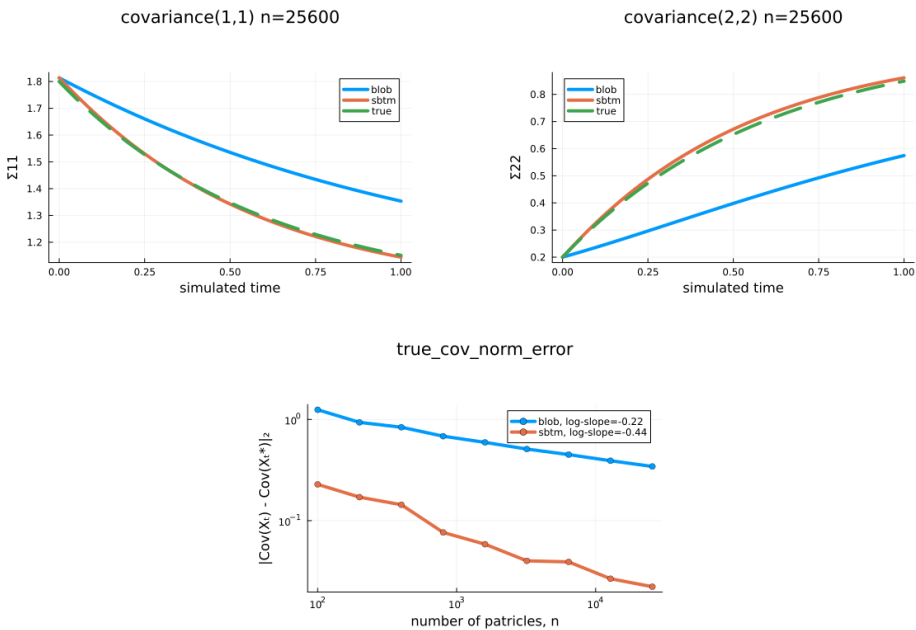


FIG. 5.4. (An anisotropic) solution to the Landau equation with Maxwellian kernel in dimension $d=10$. Even $n=25600$ particles are not enough for the blob method to match the true covariance trajectory, whereas SBTM gives a very good prediction.

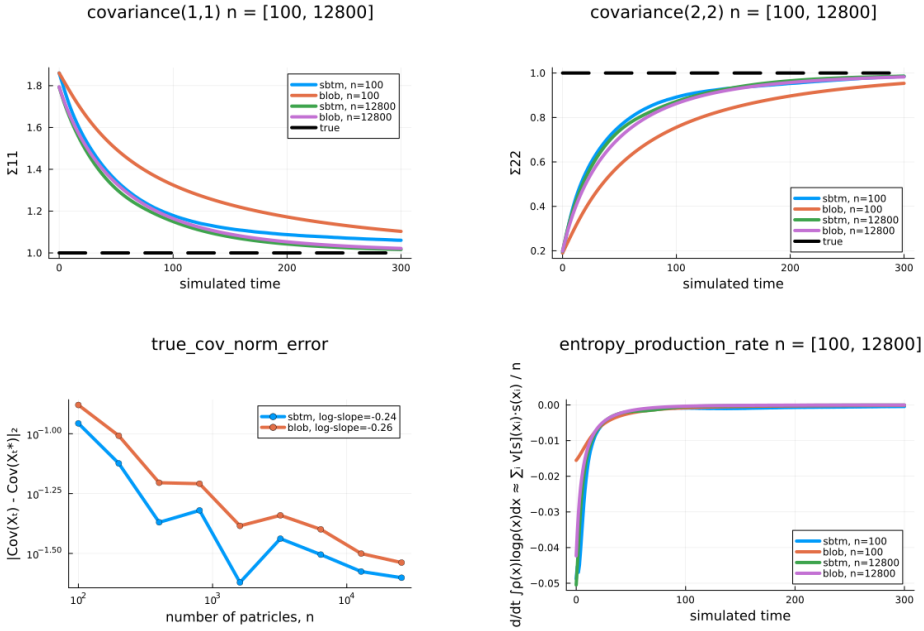


FIG. 5.5. (An anisotropic) solution to the Landau equation with Coulomb kernel in dimension $d=3$. SBTM achieves better accuracy with $n=100$ particles in terms of matching covariance and entropy decay rate of the simulations with $n=12800$ particles.

initial data. Without loss of generality, the covariance of the initial data can be taken to be diagonal. Indeed, since the covariance is positive semi-definite, there exists an orthonormal choice of basis in which covariance is diagonal. Furthermore, since the equilibrium distribution of the Landau equation is normal, it is natural to take the initial distribution to be normal with a diagonal covariance. This leads to the following example.

EXAMPLE 5.2 (Anisotropic initial condition in dimension $d=3,10$). We take the initial distribution to be normal with covariance

$$\Sigma_{i,j} = \delta_{i,j} \sigma_i, \quad \sigma_1 = 1.8, \quad \sigma_2 = 0.2, \quad \sigma_i = 1, \quad i = 3, \dots, d,$$

so that the trace is equal to d , same as in the BKW example. The evolution of the covariance matrix is

$$\begin{aligned} \Sigma_{i,j}^*(t) &= \Sigma_{i,j}^*(\infty) - (\Sigma_{i,j}^*(\infty) - \Sigma_{i,j}^*(0))e^{-4dt}, \\ \Sigma_{i,j}^*(\infty) &= \frac{2E}{d} \delta_{i,j}, \quad 2E = \text{tr}(\Sigma(0)). \end{aligned}$$

We use $\Delta t = 0.01$, and $t_{end} = 4$.

Since the true solution is not known, we cannot plot the L2 error or the score error. However, there exists a closed form solution for the second moments of the true solution, which we can compare to the second moments of the numerical solution. Figure 5.3 shows the covariance trajectories, with the dashed line denoting the analytic solution, and the Frobenius norm of the difference of second moments. The fourth subfigure is

the estimated entropy production/decay rate

$$\frac{1}{n} \sum_{i=1}^n s_t(X_i) \cdot v_t(X_i).$$

The covariance and entropy rate plots use $n=100$ and $n=12800$ particles. SBTM matches covariance well even with $n=100$ particles, while the blob method needs a lot more particles to achieve the same accuracy. Using the entropy decay trajectories with $n=12800$ as a reference, SBTM with $n=100$ matches it well, unlike the blob method.

Although the (physically relevant) Landau operator resides in three-dimensional velocity space, when it is coupled with the full Vlasov operator the ambient space is six dimension. This high dimensional problem causes a lot of numerical difficulties for the traditional particle method. In the recent work [2], the blob method is generalized to the Vlasov-Landau equation. Due to the complexity of the method, the highest dimension treated there is 1D2V (one dimension in the physical space and two dimension in the velocity space). For this reason, we also compare the blob method and SBTM in dimension $d=10$. The purpose is to examine their performance for higher dimensional problem. It is evident from Figure 5.4 that even $n=25600$ particles are not enough for the blob method to match the true covariance trajectory, whereas SBTM gives a very good prediction.

5.3. Landau equation with Coulomb kernel.

EXAMPLE 5.3 (Anisotropic initial condition in dimension $d=3$). *We take the initial distribution to be normal with covariance*

$$\Sigma_{i,j} = \delta_{i,j} \sigma_i, \quad \sigma_1 = 1.8, \quad \sigma_2 = 0.2, \quad \sigma_3 = 1,$$

so that the trace is 3, and use the Coulomb collision kernel $A(z) = \frac{1}{|z|^3}(|z|^2 I_d - z \otimes z)$. We choose $\Delta t = 1$, and $t_{\text{end}} = 300$. We do not observe any instabilities with such a large time step.

While the true covariance and entropy decay rate trajectories are unknown as well as all other statistics of the true solution, if the trajectories of the SBTM and the blob method with large n coincide, that trajectory is likely the correct one and serves as the reference for the trajectories with low n . In Figure 5.5 the first two plots are the diagonal covariance entries $\Sigma_{1,1}$ and $\Sigma_{2,2}$, with the black dashed line being the covariance of the steady state. The third plot is the Frobenius norm of the difference between the steady-state second moment matrix I_3 and the second moment matrix of the particle solution at $t = t_{\text{end}}$. The fourth plot is the estimated entropy decay rate

$$\frac{1}{n} \sum_{i=1}^n s_t(X_i) \cdot v_t(X_i).$$

The horizontal axis of first covariance and entropy rate plots is time. SBTM achieves better accuracy with $n=100$ particles in terms of matching covariance and entropy decay rate of the simulations with $n=12800$ particles.

5.4. Runtime. The runtime of the score estimation routine in SBTM is linear in the number of particles, which is the computational bottleneck. Additionally, it is easy to use the GPU to speed up the neural network training using off-the-shelf libraries, such as `Flux.jl` in `Julia`. We did this for the heat equation, where we observed a roughly $O(n^{0.6})$ scaling of the computational time on the GPU for the SBTM method.

The blob method has quadratic asymptotic complexity. It does not benefit as much from parallelization as SBTM, and has a quadratic computational time scaling.

Figure 5.6 shows the timings of the two methods as a function of the number of particles n , on the CPU and GPU. The GPU timing is for the heat equation but the score estimation routine is the bottleneck for all the equations we have tested, so most of the computational time is score estimation.

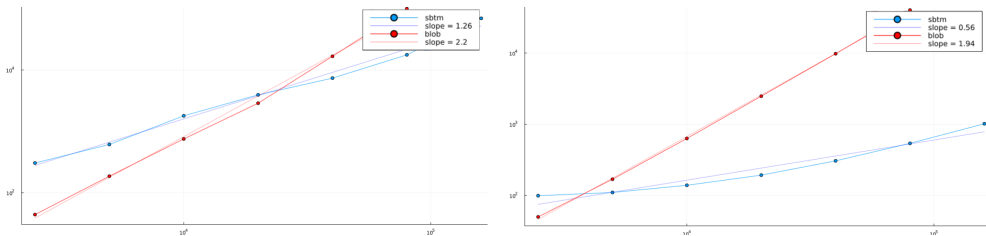


FIG. 5.6. Computational time in seconds (vertical) vs number of particles n (horizontal), on the CPU with 10 threads (left) and GPU (right) log-log scale.

6. Conclusion

In this paper, we proposed a particle method for the Fokker-Planck-Landau equation using the score-based-transport-modeling (SBTM). Compared with the traditional blob method [5], where only one parameter ε (regularization parameter for the entropy) is used to approximate the score, SBTM uses the neural network to approximate the score. We also obtained a theoretical guarantee of the method by showing that matching the score of the approximate solution is enough to recover the true solution to the Landau equation with Maxwellian molecules. Through a series of numerical examples, we demonstrated that SBTM has the following advantages over the blob method:

- (1) offers better accuracy especially for anisotropic solutions;
- (2) offers better accuracy in the sparse particles regime, e.g., high dimension or small number of particles in moderate dimension;
- (3) linear in the number of particles runtime scaling for score approximation;
- (4) easy parallelization on the GPU using off-the-shelf ML libraries, further reducing the runtime.

The main disadvantage of the SBTM method is the large number of hyperparameters, including neural network architecture, learning rate, and the number of gradient descent steps. The adaptive training strategy outlined in Section 5.1 can be more forgiving to a bad choice of hyperparameters.

Regarding the ongoing and future work: on the theoretical side, we are considering the extension of Theorem 4.2 to other values of γ , with the Coulomb kernel being of particular interest. On the numerical side, based on our preliminary analysis, making the number of gradient descent steps adaptive can greatly improve accuracy and reduce runtime. Lastly, SBTM can be used as a kernel to solve the full Vlasov-Landau Equation (1) in a similar vein as in [2]. Deriving any theoretical guarantees for this method and performing numerical experiments is an important direction for future work.

Acknowledgement. Z. Wang is partially supported by the National Key R&D Program of China, Project Number 2021YFA1002800, NSFC grant No.12171009, Young

Elite Scientist Sponsorship Program by China Association for Science and Technology (CAST) No. YESS20200028. The work of V. Ilin and J. Hu was partially supported by AFOSR grant FA9550-21-1-0358 and DOE grant DE-SC0023164.

Appendix. The continuity equation and its particle solution. In this review technical assumptions are omitted but can be found in [17]. Let $v_t(x) \in \mathbb{R}^d$ be a time-dependent vector field. Equation

$$\frac{\partial u_t}{\partial t} + \nabla \cdot (v_t u_t) = 0 \quad (\text{A.1})$$

is called a *continuity equation*. A weak solution of the continuity equation can be defined the same way as for any PDE but there is a simpler alternative definition. See Proposition 4.2 in [17] for a proof that the definition below is equivalent to the usual definition.

DEFINITION A.1 (Weak solution of continuity equation). *A weak solution of the continuity Equation (6) is a curve of probability distributions u_t indexed by $t \geq 0$ such that for all test functions $\phi \in C_c^\infty(\mathbb{R}^d)$ (compactly supported smooth functions), the following holds:*

$$\frac{d}{dt} \int_{\mathbb{R}^d} \phi(x) du_t(x) = \int_{\mathbb{R}^d} \nabla \phi(x) \cdot v_t(x) du_t(x).$$

The next proposition shows how to obtain a weak solution by solving a system of ODEs.

PROPOSITION A.1. *The particle solution given by the empirical measure as below*

$$u_t(x) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(t)}(x), \quad \frac{dX_i}{dt} = v_t(X_i)$$

is a weak solution of the continuity Equation (6).

Proof. Let $\phi(x)$ be a test function. Then

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}^d} \phi(x) du_t(x) &= \frac{d}{dt} \frac{1}{n} \sum_i \phi(X_i(t)) \\ &= \frac{1}{n} \sum_i \nabla \phi(X_i(t)) \cdot \frac{dX_i}{dt} \\ &= \int_{\mathbb{R}^d} \nabla \phi(x) \cdot v_t(x) du_t(x). \end{aligned}$$

□

REFERENCES

- [1] R. Alexandre, J. Liao, and C. Lin, *Some a priori estimates for the homogeneous Landau equation with soft potentials*, Kinet. Relat. Models, **8(4):617–650**, 2015. [2](#)
- [2] R. Bailo, J. Carrillo, and J. Hu, *The collisional particle-in-cell method for the Vlasov-Maxwell-Landau equations*, J. Plasma Phys., **90:905900415**, 2024. [5.2](#), [6](#)
- [3] N.M. Boffi and E. Vanden-Eijnden, *Probability flow solution of the Fokker-Planck equation*, Mach. Learn. Sci. Tech., **4(3):035012**, 2023. [1](#), [3.1](#), [4](#), [4.1](#)
- [4] F. Bolley and C. Villani, *Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities*, Ann. Fac. Sci. Toulouse Math., **14(3):331–352**, 2005. [4.2](#)

- [5] J.A. Carrillo, J. Hu, L. Wang, and J. Wu, *A particle method for the homogeneous Landau equation*, J. Comput. Phys. X, **7**:100066, 2020. [1](#), [3.1](#), [5](#), [5.1](#), [5.1](#), [6](#)
- [6] J.A. Carrillo, K. Craig, and F.S. Patacchini, *A blob method for diffusion*, Calc. Var. Part. Diff. Eqs., **58**:1–53, 2019. [1](#)
- [7] J. Carrillo, X. Feng, S. Guo, P.-E. Jabin, and Z. Wang, *Relative entropy method for particle approximation of the Landau equation for Maxwellian molecules*, arXiv preprint, [arXiv:2408.15035](#), 2024. [2](#), [2](#)
- [8] S. Chen, S. Chewi, J. Li, Y. Li, A. Salim, and A.R. Zhang, *Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions*, Int. Conf. Learn. Rep., 2023. [1](#)
- [9] Y.-C. Chen, *A tutorial on kernel density estimation and recent advances*, Biostat. Epidemiol., **1**(1):161–187, 2017. [5.1](#), [5.1](#)
- [10] P. Degond and F.-J. Mustieles, *A deterministic approximation of diffusion equations using particles*, SIAM J. Sci. Stat. Comput., **11**(2):293–310, 1990. [1](#)
- [11] A.L. Gibbs and F.E. Su, *On choosing and bounding probability metrics*, Int. Stat. Rev., **70**(3):419–435, 2002. [4.1](#)
- [12] N. Guillen and L. Silvestre, *The Landau equation does not blow up*, arXiv preprint, [arXiv:2311.09420](#), 2023. [1](#), [2](#)
- [13] Y. Huang and L. Wang, *A score-based particle method for homogeneous Landau equation*, J. Comput. Phys., **536**:114053, 2025. [1](#)
- [14] A. Hyvärinen, *Estimation of non-normalized statistical models by score matching*, J. Mach. Learn. Res., **6**(24):695–709, 2005. [3.1](#)
- [15] J. Lu, Y. Wu, and Y. Xiang, *Score-based transport modeling for mean-field Fokker-Planck equations*, J. Comput. Phys., **503**:112859, 2024. [1](#)
- [16] D. Maoutsa, S. Reich, and M. Opper, *Interacting particle solutions of Fokker-Planck equations through gradient-log-density estimation*, Entropy, **22**(8):802, 2020. [1](#), [3.1](#)
- [17] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Birkäuser, NY, **55**:58–63, 2015. [6](#)
- [18] D.W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley & Sons, 2015. [5.1](#)
- [19] Z. Shen and Z. Wang, *Entropy-dissipation informed neural network for McKean-Vlasov type PDEs*, NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems, **59227–59238**, 2023. [1](#)
- [20] Z. Shen, Z. Wang, S. Kale, A. Ribeiro, A. Karbasi, and H. Hassani, *Self-consistency of the Fokker-Planck equation*, Proc. Mach. Learn. Res., **817–841**, 2022. [1](#)
- [21] Y. Song and S. Ermon, *Generative modeling by estimating gradients of the data distribution*, Proceedings of the 33rd International Conference on Neural Information Processing Systems, **32:11918–11930**, 2019. [1](#)
- [22] Y. Song, J. Sohl-Dickstein, D.P. Kingma, A. Kumar, S. Ermon, and B. Poole, *Score-based generative modeling through stochastic differential equations*, Int. Conf. Learn. Rep., 2021. [1](#)
- [23] C. Villani, *Decrease of the Fisher information for solutions of the spatially homogeneous Landau equation with Maxwellian molecules*, Math. Models Meth. Appl. Sci., **10**(02):153–161, 2000. [1](#)
- [24] C. Villani, *A review of mathematical topics in collisional kinetic theory*, in S. Friedlander and D. Serre (eds.), Handbook of Math. Fluid Mech. Vol. 1, North-Holland, **71–74**, 2002. [1](#)
- [25] C. Villani, *Topics in Optimal Transport*, Graduate Studies in Mathematics, Amer. Math. Soc., **58**, 2003. [2](#)
- [26] C. Villani, *On the spatially homogeneous Landau equation for Maxwellian molecules*, Math. Models Meth. Appl. Sci., **8**(06):957–983, 1998. [1](#), [2](#), [2](#)