Article

Not peer-reviewed version

Analytical and Geometric Foundations and Modern Applications of Kinetic Equations and Optimal Transport

Cécile Barbachoux and Joseph Kouneiher *

Posted Date: 31 March 2025

doi: 10.20944/preprints202503.2313.v1

Keywords: kinetic theory; Boltzmann equation; hypercoercivity; entropy dissipation; optimal transport; Wasserstein geometry; Ricci curvature; Vlasov--Poisson system; Fokker-Planck equation

Article

Analytical and Geometric Foundations and Modern Applications of Kinetic Equations and Optimal Transport

Cécile Barbachoux †,‡ and Joseph Kouneiher *,†,‡

Côte d'Azur University; cecile.barbachoux@univ-cotedazur.fr
* Correspondence: joseph.kouneiher@univ-cotedazur.fr
† 83300 Draguignan-France: Côte d'Azur University-INSPE.
‡ These authors contributed equally to this work.

Abstract: We develop a unified analytical framework linking kinetic theory, optimal transport, and entropy dissipation through the lens of hypocoercivity. Centered on the Boltzmann and Fokker–Planck equations, we analyze the emergence of macroscopic irreversibility from time-reversible dynamics via entropy methods, functional inequalities, and commutator estimates. The hypercoercivity approach provides sharp exponential convergence rates under minimal regularity, resolving degeneracies in kinetic operators through geometric control. We extend this framework to the study of hydrodynamic limits, collisional relaxation in magnetized plasmas, and the Vlasov–Poisson system for self-gravitating matter. Additionally, we explore connections with high-dimensional data analysis, where Wasserstein gradient flows, entropic regularization, and kinetic Langevin dynamics underpin modern generative and sampling algorithms. Our results highlight entropy as a structural and variational tool across both physical and algorithmic domains.

Keywords: kinetic theory; Boltzmann equation; hypercoercivity; entropy dissipation; optimal transport; Wasserstein geometry; Ricci curvature; Vlasov–Poisson system; Fokker–Planck equation

---

1. Introduction

The theory of kinetic equations provides a powerful analytical framework for describing the statistical evolution of large systems of interacting particles. Central to this framework is the Boltzmann equation, which captures the interplay between transport, collisions, and relaxation to thermodynamic equilibrium. The mathematical analysis of such equations presents profound challenges due to the nonlinearity, high dimensionality, and degeneracies present in the operators involved. Among the most significant achievements in this domain is the development of the hypocoercivity method, which rigorously quantifies the convergence to equilibrium despite the lack of uniform ellipticity. This method, introduced by Villani and further developed by Hérau, Mouhot, and others, couples entropy dissipation techniques with commutator structures and geometric control to recover coercivity in degenerate kinetic settings.

Optimal transport theory, originating in the work of Monge and Kantorovich, has recently emerged as a unifying geometric framework for understanding a wide class of dissipative and diffusive phenomena. The introduction of the Wasserstein space of probability measures as a metric space endowed with Riemannian-like structure enables a variational interpretation of many kinetic and diffusion equations. In particular, the Fokker–Planck equation can be viewed as a gradient flow of the relative entropy functional with respect to the Wasserstein metric. This geometric viewpoint reveals deep connections between curvature, functional inequalities (e.g., logarithmic Sobolev, HWI), and the rate of convergence to equilibrium.

The synthesis of hypocoercivity and optimal transport has led to significant advances in the rigorous analysis of both linear and nonlinear kinetic models. These include the derivation of exponential decay rates, propagation of regularity, stability of steady states, and quantitative hydrodynamic

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

limits. Applications span a wide range of physical systems, from collisional relaxation in magnetized plasmas and compressible flows to the long-time behavior of self-gravitating systems modeled by the Vlasov–Poisson equation.

In recent years, the techniques developed in kinetic theory and optimal transport have also found profound applications beyond traditional physical systems. Notably, in the context of data science and machine learning, the geometry of the space of probability measures, the analysis of Wasserstein gradient flows, and the structure of entropy functionals have become central to modern generative models, variational inference, and sampling algorithms. Score-based diffusion models, underdamped Langevin dynamics, and entropic regularized optimal transport (e.g., Sinkhorn distances) are now widely employed in high-dimensional statistical learning. These methods reflect, at a computational level, the same mathematical structures–entropy decay, functional inequalities, convergence in metric measure spaces–that underlie kinetic relaxation.

This paper develops a unified and rigorous perspective on these interrelated themes. We begin by revisiting the foundational aspects of the Boltzmann equation, entropy dissipation, and the H-theorem. We then present the framework of hypocoercivity in both linear and nonlinear settings, highlighting its geometric and analytical underpinnings. The role of functional inequalities, commutator estimates, and hypoellipticity is emphasized throughout. Building on this, we explore connections with optimal transport and the geometry of the Wasserstein space, with special attention to Ricci curvature lower bounds and convexity of entropy.

The latter sections of the paper are devoted to applications: we study the relaxation behavior of plasmas under external magnetic fields, the derivation of fluid models from kinetic equations, the dynamics of self-gravitating astrophysical systems, and the implementation of kinetic-inspired algorithms in data science. Throughout, we stress the conceptual role of entropy as both a physical observable and a variational structure, linking microdynamics, macrodynamics, and probabilistic learning.

Historical and Modern Developments of Optimal Transport

The history of optimal transport theory can be traced back to multiple independent discoveries, evolving through different mathematical frameworks over centuries. This text provides an overview of its foundational contributors and the subsequent evolution of the field.

The first formulation of the optimal transport problem was introduced by *Gaspard Monge* in 1781, in his *Mémoire sur la théorie des déblais et des remblais* [1]. Monge's problem involved minimizing transportation costs when moving materials from one place to another. His formulation sought a deterministic optimal coupling that would assign each unit of material to a specific destination, minimizing the total cost based on distance.

Monge's geometric intuition led to key mathematical insights, such as transport occurring along orthogonal straight lines to certain surfaces, leading to discoveries in differential geometry. However, his mathematical treatment lacked formal rigor by modern standards.

Monge's ideas resurfaced much later in the 1938 work of *Leonid Kantorovich*, a Soviet mathematician and economist, who reformulated the problem in the language of linear programming [2]. He introduced the *Kantorovich relaxation*, allowing mass to be split and transported probabilistically rather than deterministically.

Kantorovich also developed duality theory, which became fundamental in solving transport problems. His work extended beyond mathematics into economics, leading to his Nobel Prize in Economics (1975) for contributions to the theory of resource allocation. A key contribution to optimal transport was the definition of the *Kantorovich–Rubinstein distance*, a metric that measures the cost of transporting one probability measure into another [3].

Throughout the mid-to-late 20th century, statisticians and probabilists expanded on Kantorovich's ideas, particularly in probability theory and functional analysis. In the 1970s, *Dobrushin* applied optimal transport distances to study interacting particle systems [5] . *Hiroshi Tanaka* used these techniques in kinetic theory, particularly in understanding variants of the Boltzmann equation [6].

 doi:10.20944/preprints202503.2313.v1

By the 1980s, three independent research directions emerged that reshaped the field: *John Mather* (Dynamical Systems) connected action-minimizing curves in Lagrangian mechanics with optimal transport problems [7]; *Yann Brenier* (Fluid Mechanics and PDEs) established links between OT and incompressible fluid mechanics, particularly via the Monge–Ampère equation and convex analysis [8]; and *Mike Cullen* (Meteorology) showed that semi-geostrophic equations in meteorology could be reinterpreted using optimal transport principles [9].

A major turning point came in the early 2000s with the groundbreaking work of *Cédric Villani*, who systematically unified the field and extended its applications across geometry, analysis, and physics. His two monographs, *Topics in Optimal Transportation* (2003) and *Optimal Transport: Old and New* (2009) [3,4], became foundational texts, synthesizing decades of fragmented work and establishing a coherent theoretical framework.

Villani's work, often in collaboration with researchers such as Léonard, Ambrosio, McCann, Otto, and others, led to the geometrization of probability spaces using optimal transport. He helped formalize the *Wasserstein geometry* on the space of probability measures, enabling a differential structure akin to Riemannian geometry. This gave rise to: gradient flows in the Wasserstein space (developed notably by Felix Otto and Ambrosio–Gigli–Savaré) [10,11]; new insights into Ricci curvature bounds on metric measure spaces via the Lott–Sturm–Villani theory [12]; and applications to entropy, diffusion, and functional inequalities (e.g., Talagrand, HWI inequalities) [4]. These developments had powerful implications in geometric analysis, particularly in understanding spaces with lower bounds on Ricci curvature and the analysis of heat flow in non-smooth settings.

In the 21st century, optimal transport has become a highly interdisciplinary field, with vibrant applications in machine learning and data science–particularly in generative models (e.g., Wasserstein GANs) [13], domain adaptation, clustering, and distributional learning; in image processing and computer vision, including color transfer, shape matching, and texture synthesis [14]; in economics, especially in matching theory and income inequality metrics; in statistics, for defining distances between distributions in high dimensions [15]; and in quantum physics, statistical mechanics, and density functional theory.

Recent advances have also explored unbalanced transport, where mass is allowed to be created or destroyed (e.g., Chizat, Peyré, Schmitzer) [18]; entropy-regularized OT, making computation feasible at large scales (e.g., Sinkhorn distances) [17]; discrete OT for graphs and networks; dynamic formulations (Benamou–Brenier), leading to efficient numerical methods [19]; and barycenters in Wasserstein space, with applications in image averaging and consensus learning [20].

The evolution of optimal transport theory showcases the power of mathematical abstraction to transcend disciplinary boundaries. From Monge's geometric intuition to Kantorovich's probabilistic reformulation, and culminating in the modern theory shaped by Villani and his collaborators, optimal transport has become a central tool in mathematics and beyond. Its current growth is fueled by its unifying nature, geometric depth, and computational versatility, with active research directions still unfolding across mathematics, computer science, physics, and the social sciences.

This paper explores the foundational issues underlying these theories, with an emphasis on stability, entropy methods, hypoellipticity, and geometric connections. We will systematically analyze the key principles and challenges associated with each domain, shedding light on their intersections and mathematical richness.

2. Boltzmann Breakthrough: Kinetic Theory, Optimal Transport, and Entropy

Mathematics plays a crucial role in describing the fundamental processes governing natural and physical phenomena. Among the various branches of applied mathematics, *kinetic theory* and *optimal transport* have emerged as essential tools in understanding how particles and probability distributions evolve over time. Kinetic equations describe the statistical behavior of particle systems, either with or without collisions, while optimal transport theory provides a powerful framework for studying the

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

movement of mass in the most efficient manner. Both areas have profound implications, from plasma physics and fluid mechanics to geometry and functional analysis.

The mathematical study of kinetic equations dates back to Ludwig Boltzmann's pioneering work in the 19th century, leading to the well-known *Boltzmann equation* that models gas dynamics:

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = Q(f, f), \tag{1}$$

where $f(t, x, v)$ is the *distribution function*, describing the probability density of particles at time t, position $x \in \mathbb{R}^d$, and velocity $v \in \mathbb{R}^d$.

The term $Q(f, f)$ is the *collision operator*, which accounts for the change in velocity distribution due to interactions between particles. It encodes the fundamental mechanism by which a gas approaches thermal equilibrium. The Boltzmann equation provides a statistical description of a system with many interacting particles, bridging the microscopic laws of physics with macroscopic thermodynamic behavior. In its classical form for hard spheres:

$$Q(f, f)(v) = \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B(|v - v_*|, \cos\theta) \left[f(v') f(v'_*) - f(v) f(v_*) \right] d\sigma dv_*, \tag{2}$$

where (v, v_*) and (v', v'_*) denote the pre- and post-collision velocities, related via the scattering laws, and B is the collision kernel depending on relative velocity and scattering angle.

This equation plays a central role in kinetic theory, as it models how particle collisions influence the macroscopic behavior of a gas. It encapsulates the transition from microscopic Newtonian interactions to emergent thermodynamic laws.

The Boltzmann equation marks a significant conceptual shift in the understanding of physical systems. Historically, classical mechanics provided deterministic descriptions of particle motion, governed by Newton's laws. In contrast, kinetic theory introduces a statistical perspective, acknowledging the impracticality of tracking every individual particle in a large system. This shift from a deterministic to a probabilistic framework highlights the deep epistemological divide between microscopic mechanics and macroscopic thermodynamics.

The function $f(t, x, v)$ encapsulates our knowledge of a system not in terms of precise trajectories but in terms of probability distributions. This probabilistic description aligns with the broader conceptual transition in physics from classical determinism to statistical and quantum interpretations. The use of distribution functions reflects an epistemological necessity: our inability to resolve individual particle positions and velocities necessitates a coarse-grained, statistical approach.

Furthermore, the introduction of the *collision operator* $Q(f, f)$ represents an abstraction of microscopic interactions, reducing complex many-body dynamics into an effective statistical mechanism. This reduction raises questions about the emergent nature of macroscopic laws: how do local, microscopic interactions give rise to global, thermodynamic behavior? The principle of *entropy increase*, embedded in Boltzmann's H-theorem, illustrates how irreversibility emerges from time-reversible microscopic laws. This paradox, deeply connected to Loschmidt's and Zermelo's objections to Boltzmann's theory, remains a foundational issue in the philosophy of physics.

A key insight is encoded in Boltzmann's celebrated *H-theorem*, which introduces the *entropy functional*

$$\mathcal{H}(f)(t) := \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(t, x, v) \log f(t, x, v) \, dv dx, \tag{3}$$

and asserts that its time derivative satisfies

$$\frac{d}{dt} \mathcal{H}(f)(t) \leq 0, \tag{4}$$

with equality only at equilibrium (e.g., Maxwellian distribution). This inequality reflects the second law of thermodynamics: entropy does not decrease.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

This result–despite being derived from time-reversible microscopic dynamics–predicts the irreversible trend toward equilibrium, generating tension with the reversibility of Newtonian mechanics (as noted in Loschmidt's paradox and Zermelo's recurrence objection). These philosophical challenges remain foundational in statistical physics [26,27].

Moreover, kinetic theory serves as a bridge between various mathematical and physical domains. It connects functional analysis, measure theory, and PDE theory with physical concepts such as equilibrium, fluctuations, and dissipation. The Boltzmann equation is a nonlinear integro-differential equation, and its study has led to significant advances in the theory of PDEs, particularly in hypoellipticity and hypocoercivity [28,29].

Finally, the Boltzmann equation and its generalizations continue to inform modern research in non-equilibrium statistical mechanics, stochastic processes, and even quantum kinetic theory. The conceptual and foundational challenges it poses, such as the justification of the molecular chaos hypothesis, the nature of entropy, and the emergence of macroscopic irreversibility, remain at the heart of ongoing discussions in both mathematical physics and the philosophy of science.

More recently, the *Vlasov equation*

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + F[f] \cdot \nabla_v f = 0 \tag{5}$$

has been used to describe large-scale astrophysical and plasma systems where collisions are negligible. Here, $F[f]$ is a self-consistent force field derived from f, e.g., via Poisson or Maxwell equations. In this collisionless regime, questions about Landau damping, plasma echo, and long-time stability dominate [30].

Parallel to kinetic theory, *optimal transport* has provided deep insights into the geometry of probability distributions and functional inequalities. Optimal transport theory now connects Ricci curvature [12], statistical mechanics and entropy [10], partial differential equations and diffusion via Wasserstein geometry.

The Wasserstein distance W_2 between two probability densities μ, ν is defined as

$$W_2^2(\mu, \nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \, d\gamma(x,y), \tag{6}$$

where $\Pi(\mu, \nu)$ is the set of couplings with marginals μ and ν. This defines a geodesic metric on the space of probability measures with finite second moments.

By bridging these fields, optimal transport offers novel perspectives on fundamental mathematical problems.

3. Entropy and the H-Theorem

A key feature of the Boltzmann equation is its deep connection to *entropy*. The *Boltzmann entropy*, denoted by $H(f)$, is defined as:

$$H(f) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x,v) \log f(x,v) \, dv dx. \tag{7}$$

This function measures the disorder in the system and plays a crucial role in thermodynamic laws. Ludwig Boltzmann introduced the celebrated *H-theorem*, which states that entropy increases over time:

$$\frac{dH}{dt} \leq 0. \tag{8}$$

More precisely, the entropy dissipation rate $D(f)$ is defined by:

$$D(f) := -\frac{d}{dt} H(f) = -\int Q(f,f) \log f \, dv, \tag{9}$$

which is non-negative, i.e., $D(f) \geq 0$. This provides a statistical explanation for the second law of thermodynamics: a closed system will evolve irreversibly toward a state of maximum entropy, corresponding to *thermal equilibrium*.

To understand the mechanism behind entropy production, we note that collisions in the Boltzmann equation drive the system towards a *Maxwellian equilibrium*:

$$f_\infty(v) = \frac{\rho}{(2\pi T)^{d/2}} \exp\left(-\frac{|v-u|^2}{2T}\right), \tag{10}$$

where ρ is the density, T the temperature, and u the mean velocity of the gas. The entropy of this Maxwellian distribution is maximized, which explains why physical systems naturally evolve toward this state.

3.1. Cercignani's Conjecture and Entropy Dissipation

While the H-theorem establishes that entropy increases, it does not provide an explicit rate of convergence to equilibrium. *Cercignani's conjecture* [31] refines this understanding by proposing a quantitative relationship between entropy dissipation and deviation from equilibrium. The conjecture states:

$$D(f) \geq C(H(f) - H(f_\infty)), \tag{11}$$

where $C > 0$ is a constant depending on the collision kernel and physical parameters. This inequality suggests that the closer the system is to equilibrium, the slower the entropy dissipation, leading to an explicit control of the convergence rate.

In a precise sense, $D(f)$ quantifies the rate at which entropy is produced in a system described by the Boltzmann equation. Conceptually, it measures how fast the system evolves towards equilibrium by accounting for the effects of collisions on the distribution function. It is a key quantity in proving stability and convergence results, with deep connections to *functional inequalities*, such as logarithmic Sobolev inequalities and the spectral gap.

Thus, the entropy dissipation rate $D(f)$ serves as a fundamental bridge between microscopic dynamics (collisions) and macroscopic thermodynamic behavior (irreversibility and equilibrium). In kinetic theory, collisions redistribute velocities, and $D(f)$ reflects the effectiveness of this redistribution. As per the H-theorem, we have:

$$\frac{dH(f)}{dt} = -D(f) \leq 0. \tag{12}$$

Entropy production arises from the redistribution of particles in velocity space. The stronger the collisions (i.e., the more mixing occurs), the greater the entropy dissipation, meaning the system reaches equilibrium faster.

Cercignani's conjecture, restated as:

$$D(f) \geq C(H(f) - H(f_\infty)), \tag{13}$$

connects entropy dissipation to the distance from equilibrium. It emphasizes the stabilizing role of collisions in driving the system toward maximum entropy.

The concept of entropy dissipation extends beyond the Boltzmann equation into broader thermodynamic contexts. It represents the rate at which a system loses free energy due to internal interactions. In fluid dynamics, for example, entropy dissipation is analogous to *viscous dissipation*, where kinetic energy is irreversibly converted into heat.

Cercignani's conjecture remained an open problem for many years and is known to be *not* universally true in its original form. However, significant progress was made by *Villani* and *Toscani* [26, 32], who established modified versions of the conjecture. In particular, they proved:

$$D(f) \geq \lambda\,\Phi(H(f) - H(f_\infty)), \tag{14}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

for some convex function Φ, under suitable regularity and moment assumptions. These results provided a rigorous framework for quantifying entropy dissipation and convergence rates in kinetic theory.

4. Hypercoercivity: Resolving Degeneracy and Ensuring Convergence

One of the central mathematical challenges in analyzing the long-time behavior of kinetic equations–such as the Boltzmann or linear Fokker-Planck equations–is the issue of *degeneracy* in the collision or diffusion operator. Specifically, in the Boltzmann equation,

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = Q(f, f),$$

the *collision operator* $Q(f, f)$ acts only in the velocity variable v and leaves the spatial variable x untouched. This degeneracy obstructs the direct application of standard coercivity arguments (such as Poincaré or spectral gap inequalities) in the full phase space (x, v).

In particular, the linearized Boltzmann equation around a Maxwellian equilibrium $f_\infty(v)$ is typically written as:

$$\partial_t h + v \cdot \nabla_x h = \mathcal{L}h,$$

where $h := f - f_\infty$, and $\mathcal{L}$ is the linearized collision operator. While $\mathcal{L}$ is coercive in v (modulo kernel), the transport term $v \cdot \nabla_x$ introduces oscillations and mixing that are not controlled directly by $\mathcal{L}$.

To overcome this, *Cédric Villani* introduced the method of *hypercoercivity* [28], a general framework designed to handle this type of degeneracy and establish quantitative exponential convergence rates toward equilibrium.

Functional Setting and Degeneracy

The degeneracy manifests in the lack of full coercivity of the generator of the kinetic semigroup. Consider a Hilbert space $\mathcal{H}$, such as $L^2(\mathbb{T}^d_x \times \mathbb{R}^d_v, \mu^{-1} dxdv)$, where $\mu(v) = e^{-|v|^2/2}$ is the Gaussian or Maxwellian weight. Define the evolution operator:

$$\mathcal{A} := -v \cdot \nabla_x + \mathcal{L}.$$

The operator $\mathcal{L}$ typically satisfies:

$$\langle \mathcal{L}h, h \rangle \leq -\lambda \|h^\perp\|^2,$$

where $h^\perp$ is the projection of h orthogonal to the kernel of $\mathcal{L}$ (i.e., macroscopic modes). However, $\mathcal{A}$ is not coercive in $\mathcal{H}$ due to the transport term, and in fact, it may not even be sectorial.

Villani's Hypercoercivity Method

Villani's key insight was to modify the energy functional to include cross-derivative terms that couple x and v regularities, in order to exploit *commutator structure* and transfer the velocity dissipation to spatial variables.

Let us define the modified energy functional:

$$\mathcal{E}(h) := \|h\|^2 + \alpha \langle \nabla_x h, \nabla_v h \rangle + \beta \|\nabla_x h\|^2,$$

with appropriately chosen $\alpha, \beta > 0$. Then one shows that:

$$\frac{d}{dt} \mathcal{E}(h(t)) \leq -\lambda \mathcal{E}(h(t)),$$

for some explicit $\lambda > 0$, leading to:

$$\mathcal{E}(h(t)) \leq \mathcal{E}(h_0) e^{-\lambda t}.$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

This decay estimate implies that:

$$\|h(t)\|_{L^2} \leq Ce^{-\lambda t}\|h(0)\|_{H^1}, \tag{15}$$

which demonstrates exponential convergence to equilibrium in a weighted L^2 norm.

Geometric and Analytical Structure

The success of the hypercoercivity method relies on three interconnected ingredients:

- *Entropy dissipation*: The collision operator provides dissipation in the v variable, which controls the non-equilibrium modes.
- *Hypoellipticity and regularity transfer*: Inspired by Hörmander's theory [33], certain commutators between the transport operator and the collision operator generate smoothing effects in x.
- *Commutator estimates*: The key to propagating dissipation from velocity to spatial variables involves bounding expressions like $[\nabla_x, \mathcal{L}^{-1} v \cdot \nabla_x]$ or higher-order mixed derivatives.

The convergence result can be interpreted as a non-symmetric analogue of coercivity: although the generator is not coercive in the standard sense, the flow generated by it dissipates energy due to the interaction between the dissipative and conservative directions.

Abstract Hypocoercivity Theorem (Villani)

Let $A = B + C$ on a Hilbert space $\mathcal{H}$, with B symmetric and dissipative, C antisymmetric (e.g., $C = v \cdot \nabla_x$). Under certain bracket conditions:

$$\exists\, n \in \mathbb{N} \text{ such that } \mathrm{Lie}^n(B, C) \supset \mathrm{Id},$$

and assuming that B has a spectral gap, then A generates a semigroup satisfying:

$$\|e^{tA} f - f_\infty\| \leq Ce^{-\lambda t}\|f - f_\infty\|.$$

This gives exponential decay toward equilibrium with explicit control on the rate λ.

Applications and Extensions

This theory has been successfully applied to a wide class of kinetic models:

- Linearized Boltzmann and Landau equations with periodic or confining domains [30,36].
- Kinetic Fokker–Planck equations with external potentials [34,35].
- Quantum and semi-classical limits of collisional kinetic models.

Villani's hypercoercivity program provides a unified functional analytic framework for proving convergence to equilibrium in degenerate, non-symmetric PDEs where classical coercivity fails. It combines tools from semigroup theory, PDEs, microlocal analysis, and differential geometry.

$$\|f(t, x, v) - f_\infty\|_{L^2_\mu} \leq Ce^{-\lambda t}\|f(0, x, v) - f_\infty\|_{H^1_\mu}, \tag{16}$$

where $\mu(v) = e^{-|v|^2/2}$ is the Maxwellian weight.

5. Wasserstein Geometry and Villani's Contributions to Optimal Transport

Villani's work, often in collaboration with researchers such as Léonard, Ambrosio, McCann, Otto, and others, led to the geometrization of probability spaces using optimal transport. He helped formalize the *Wasserstein geometry* on the space of probability measures, enabling a differential structure akin to Riemannian geometry.

 doi:10.20944/preprints202503.2313.v1

Let $\mathcal{P}_2(M)$ be the space of Borel probability measures on a Riemannian manifold M with finite second moment. The 2-*Wasserstein distance* between $\mu, \nu \in \mathcal{P}_2(M)$ is defined as:

$$W_2(\mu,\nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{M\times M} d(x,y)^2 \, d\pi(x,y) \right)^{1/2},$$

where $\Pi(\mu,\nu)$ is the set of transport plans, i.e., Borel probability measures on $M \times M$ with marginals μ and ν.

This metric endows $\mathcal{P}_2(M)$ with a geodesic structure: there exists a constant-speed geodesic $(\mu_t)_{t\in[0,1]}$ between any two measures μ_0 and μ_1. This structure is key to formulating *displacement convexity*, an idea introduced by McCann [21].

In the early 2000s, Felix Otto observed that the heat equation on $\mathbb{R}^n$,

$$\partial_t \rho = \Delta \rho,$$

can be viewed as the *gradient flow* of the entropy functional in the space $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ [10]. That is, the dynamics of ρ is a steepest descent of the Boltzmann entropy:

$$\mathcal{H}(\rho) := \int_{\mathbb{R}^n} \rho(x) \log \rho(x) \, dx.$$

This interpretation led to a formal Riemannian structure on $\mathcal{P}_2$, defined rigorously through the dynamic formulation of W_2 by Benamou and Brenier [19]:

$$W_2^2(\mu_0,\mu_1) = \inf_{\rho_t, v_t} \left\{ \int_0^1 \int_{\mathbb{R}^n} \rho_t(x) \|v_t(x)\|^2 dx \, dt \;\middle|\; \partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0 \right\}.$$

This geometrization was rigorously developed by Ambrosio, Gigli, and Savaré [11], who created a full theory of gradient flows in metric spaces. Their work introduced *evolution variational inequalities* (EVIs) and characterized λ-convex functionals in Wasserstein space.

Villani, together with Lott [12] and independently Sturm [22,23], extended this framework to general metric measure spaces (X, d, m) via convexity of the entropy along Wasserstein geodesics. Let

$$\mathcal{H}_m(\mu) := \int \log\left(\frac{d\mu}{dm}\right) d\mu$$

be the relative entropy w.r.t. a reference measure m. The space satisfies a curvature-dimension condition $CD(K,\infty)$ if for any geodesic $(\mu_t)_{t\in[0,1]}$ in $\mathcal{P}_2(X)$,

$$\mathcal{H}_m(\mu_t) \le (1-t)\mathcal{H}_m(\mu_0) + t\mathcal{H}_m(\mu_1) - \frac{K}{2} t(1-t) W_2^2(\mu_0,\mu_1).$$

This *Lott–Sturm–Villani theory* generalizes Ricci curvature lower bounds to singular spaces, and applies to heat flow, functional inequalities, and geometric analysis.

One of the celebrated applications of this geometric formalism is *Talagrand's inequality* [24], which asserts that for the standard Gaussian measure γ and any measure μ absolutely continuous w.r.t. γ:

$$W_2^2(\mu,\gamma) \le 2\,\mathrm{Ent}_\gamma(\mu).$$

Building on this, Otto and Villani derived the *HWI inequality* [25], interpolating between relative entropy H, Wasserstein distance W, and Fisher information I:

$$\mathrm{Ent}_\gamma(\mu) \le W_2(\mu,\gamma)\sqrt{I(\mu|\gamma)} - \frac{K}{2} W_2^2(\mu,\gamma),$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

where
$$I(\mu|\gamma) := \int \left\| \nabla \log\left(\frac{d\mu}{d\gamma}\right) \right\|^2 d\mu.$$

These inequalities imply logarithmic Sobolev inequalities, hypercontractivity, and exponential convergence to equilibrium, and they demonstrate the profound unification of geometry, analysis, and probability enabled by optimal transport.

Thus, Villani's work and its extensions reshaped entire domains of geometric analysis, particularly the theory of metric measure spaces with curvature bounds, and offered a powerful, flexible framework to study heat diffusion, entropy dissipation, and nonlinear PDEs from a variational and geometric perspective.

Commutator Estimates and Propagation of Regularity

A critical aspect of the hypercoercivity framework developed by Villani [28] is the propagation of regularity from the velocity variable v to the spatial variable x. This is achieved through a careful analysis of *commutator estimates*, which exploit the non-commutative algebra of the vector fields involved in the kinetic equation.

Consider the kinetic transport operator:

$$T := \partial_t + v \cdot \nabla_x,$$

and the linearized kinetic equation:
$$Tf = \mathcal{L}f,$$

where $\mathcal{L}$ is the linearized collision operator, which acts only in the velocity variable v. While $\mathcal{L}$ provides coercivity in v, it does not directly control $\nabla_x f$. To resolve this, Villani's insight was to consider the Lie algebra generated by the differential operators appearing in the system.

Define first-order differential operators:

$$X_0 := \partial_t + v \cdot \nabla_x, \quad X_i := \partial_{v_i}, \quad i = 1, \ldots, d.$$

Note that:
$$[X_0, X_i] = \partial_{x_i}.$$

Thus, although ∂_{x_i} is not in the original list of vector fields, it is obtained via a commutator. This aligns with the structure required by Hörmander's hypoellipticity theorem [33]: if the Lie algebra generated by a collection of vector fields spans the tangent space at each point, then the associated operator is hypoelliptic.

This commutator structure implies that regularity in the velocity variable, when combined with transport in x, induces regularity in x. This mechanism is formalized through estimates of the form:

$$\|[X_i, X_j]f\|_{L^2} \le C\|\mathcal{L}f\|_{L^2}, \tag{17}$$

where $[X_i, X_j] := X_iX_j - X_jX_i$ denotes the commutator of two operators. Since $\mathcal{L}$ provides dissipation in v, and since ∇_x can be expressed as a commutator involving ∇_v, we obtain an indirect control over the spatial derivatives of f.

This transfer of regularity is crucial for constructing coercive energy functionals, as in the hypercoercive Lyapunov framework:

$$\mathcal{E}(f) := \|f\|^2 + \alpha\langle \nabla_x f, \nabla_v f\rangle + \beta\|\nabla_x f\|^2.$$

Differentiating $\mathcal{E}(f(t))$ along the solution and applying commutator bounds (such as (17)) yields:

$$\frac{d}{dt}\mathcal{E}(f(t)) \le -\lambda\mathcal{E}(f(t)) + \text{lower-order terms},$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

which shows that the modified energy decays exponentially with time.

This process can be interpreted geometrically: the solution space is equipped with a sub-Riemannian structure, where smoothing in x is induced by the geometry of commutators between the transport and diffusion vector fields.

Explicit Convergence Rate

By combining entropy dissipation, hypoelliptic smoothing, and commutator estimates, the hypercoercivity method allows one to rigorously prove *explicit exponential convergence to equilibrium*. For a wide class of linear kinetic equations–including the linearized Boltzmann and Fokker–Planck equations–one obtains:

$$\|f(t,x,v) - f_\infty(v)\|_{L^2_\mu} \leq Ce^{-\lambda t}\|f(0,x,v) - f_\infty(v)\|_{H^1_\mu}, \tag{18}$$

where $\mu(v) = e^{-|v|^2/2}$ is the Maxwellian weight, f_∞ is the equilibrium state (often a global Maxwellian), and the constants $C, \lambda > 0$ depend on parameters such as the collisional cross-section, dimension d, and domain geometry [34,35].

This result rigorously confirms that any smooth solution of the linearized kinetic equation with periodic or confined spatial domain converges toward equilibrium at an explicitly quantifiable rate.

In the nonlinear case, such as the full Boltzmann equation with hard spheres and periodic boundary conditions, similar exponential decay results have been obtained under close-to-equilibrium assumptions via nonlinear hypocoercivity methods [30,36].

These quantitative estimates validate Boltzmann's physical intuition about the *irreversible trend toward equilibrium* and connect probabilistic entropy arguments with sharp analytic inequalities.

6. Applications and Broader Implications

The study of entropy production, hypocoercivity, and stability in kinetic equations has far-reaching consequences across mathematical physics, applied mathematics, and geometry. These methods establish rigorous bridges between the microscopic particle description of systems and their macroscopic thermodynamic behavior, enabling multiscale modeling and convergence analysis in a variety of settings.

6.1. Plasma Physics: Kinetic Relaxation in Magnetized Systems

The classical coercivity techniques fall short in the presence of degeneracies introduced by transport operators, especially when collisions act only in the velocity variable. The method of hypercoercivity, introduced by Villani [28], addresses this by coupling dissipative and conservative effects through modified energy functionals. In the context of magnetized plasmas in physics, we consider the Vlasov-Poisson-Boltzmann or Vlasov–Maxwell–Landau systems in a spatial domain $\Omega \subset \mathbb{R}^3$ with periodic boundary conditions. The evolution of the distribution function $f(t,x,v)$ for ions is governed by:

$$\partial_t f + v \cdot \nabla_x f + (E + v \times B) \cdot \nabla_v f = Q(f,f), \tag{19}$$

where:

- $E = -\nabla_x \phi$ is the electric field derived from the potential ϕ,
- B is a constant external magnetic field,
- $Q(f,f)$ is the Boltzmann collision operator.

The self-consistent potential satisfies:

$$\Delta_x \phi = \rho_f(x) - \rho_0, \quad \rho_f(x) = \int f(x,v)\, dv, \tag{20}$$

where ρ_0 is the background ion density.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

We can rewrite the equation (19) in the form

$$\partial_t f + v \cdot \nabla_x f + F[f] \cdot \nabla_v f = Q(f, f), \tag{21}$$

where $F[f]$ is the self-consistent force derived from the electric or magnetic field (e.g., $F = -\nabla_x \phi$, with ϕ solving Poisson's equation), and $Q(f, f)$ describes collisional interactions (Boltzmann or Landau).

Entropy dissipation plays a crucial role in understanding collisional relaxation in magnetized plasmas. The entropy functional:

$$\mathcal{H}(f) = \int f \log f \, dxdv, \tag{22}$$

decreases in time, and its dissipation rate governs the transition to thermal equilibrium. The hypocoercivity framework allows one to obtain exponential decay toward Maxwellian states even in the presence of magnetic field-induced degeneracies [37,38].

Commutator Structures in Magnetized Plasmas

The Lorentz force term $v \times B$ poses additional challenges, as it induces rotations in velocity space. However, using commutators such as:

$$[\partial_{v_i}, v_j \partial_{x_j}] = \delta_{ij} \partial_{x_j}, \tag{23}$$

and analyzing the Lie algebra generated by transport and collision directions, we obtain hypoelliptic smoothing and full control over all derivatives.

Numerical Simulations

Numerical schemes preserving entropy dissipation are crucial for simulating kinetic equations. We implement a spectral method for the velocity variable and a finite-volume scheme for space, preserving conservation laws and entropy decay.

Simulations show that the distribution $f(t, x, v)$ converges exponentially toward equilibrium, confirming theoretical decay rates. The effect of magnetic field intensity $|B|$ is observed: stronger fields slow spatial mixing but enhance velocity-space regularization.

Landau Operator Variant

The Landau operator is a limit of the Boltzmann operator for grazing collisions and reads:

$$Q_L(f, f) = \nabla_v \cdot \left(\int a(v - v_*) [f(v_*) \nabla_v f(v) - f(v) \nabla_{v_*} f(v_*)] dv_* \right), \tag{24}$$

where $a(z)$ is a positive semi-definite matrix depending on $|z|$. For Maxwellian molecules:

$$a(z) = |z|^{\gamma+2} \left(I - \frac{z \otimes z}{|z|^2} \right), \quad \gamma = 0. \tag{25}$$

The hypercoercivity framework extends to Landau-type operators, yielding similar exponential convergence results under modified functional settings.

Our analysis confirms that collisional relaxation in magnetized plasmas leads to exponential convergence toward Maxwellian equilibrium, with explicit decay rates depending on collision frequency and magnetic field intensity. This provides a theoretical justification for numerical observations of thermalization in magnetically confined fusion devices.

6.2. Fluid Dynamics: Hydrodynamic Limits and Dissipation

Kinetic equations serve as mesoscopic models that bridge the microscopic world of particles and the macroscopic continuum descriptions used in fluid dynamics. In particular, the Boltzmann equation provides a probabilistic description of a dilute gas, while the Euler and Navier–Stokes equations

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

govern the behavior of compressible and incompressible fluids, respectively. The connection between these descriptions is made rigorous through the study of hydrodynamic limits.

Consider the rescaled Boltzmann equation:

$$\epsilon \partial_t f + v \cdot \nabla_x f = \frac{1}{\epsilon} Q(f, f), \quad \epsilon \to 0, \tag{26}$$

where ϵ is the Knudsen number, representing the ratio of the mean free path to the macroscopic length scale. This scaling corresponds to the so-called *fluid dynamic regime*.

The formal limit $\epsilon \to 0$ leads to the
emphlocal equilibrium assumption:

$$Q(f, f) = 0 \quad \Rightarrow \quad f \approx M_{[\rho,u,T]}(v), \tag{27}$$

where $M_{[\rho,u,T]}$ is the local Maxwellian defined by

$$M_{[\rho,u,T]}(v) = \frac{\rho}{(2\pi T)^{3/2}} \exp\left(-\frac{|v-u|^2}{2T}\right), \tag{28}$$

with ρ the mass density, u the mean velocity, and T the temperature.

Integrating the Boltzmann equation against collision invariants 1, v, and $|v|^2$ yields the conservation laws for mass, momentum, and energy:

$$\partial_t \rho + \nabla_x \cdot (\rho u) = 0, \tag{29}$$

$$\partial_t (\rho u) + \nabla_x \cdot (\rho u \otimes u + pI) = 0, \tag{30}$$

$$\partial_t E + \nabla_x \cdot ((E + p)u) = 0, \tag{31}$$

where $E = \frac{1}{2}\rho|u|^2 + \frac{3}{2}\rho T$ is the total energy and $p = \rho T$ is the pressure (ideal gas law).

These equations correspond to the compressible Euler system. To derive the Navier–Stokes system, one needs a higher-order Chapman–Enskog expansion:

$$f = M + \epsilon f_1 + \epsilon^2 f_2 + \cdots, \tag{32}$$

where f_1 solves the linearized Boltzmann equation:

$$Q(M, f_1) + Q(f_1, M) = -(\partial_t M + v \cdot \nabla_x M). \tag{33}$$

This expansion allows one to derive constitutive relations for the stress tensor σ and heat flux q:

$$\sigma = \mu(\nabla_x u + \nabla_x u^T - \tfrac{2}{3} \operatorname{div} u\, I), \tag{34}$$

$$q = -\kappa \nabla_x T, \tag{35}$$

leading to the compressible Navier–Stokes equations.

From a mathematical standpoint, passing to the limit $\epsilon \to 0$ rigorously requires compactness, uniform estimates, and control of entropy production. The entropy inequality:

$$\frac{d}{dt} \int f \log f \, dx dv \leq 0, \tag{36}$$

combined with suitable bounds on moments and dissipation,

$$\int_0^T \int \frac{1}{\epsilon} D(f) \, dx dt < \infty, \tag{37}$$

allows the use of compactness tools (e.g., velocity averaging lemmas) to establish weak convergence.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

Hypercoercivity and Near-Equilibrium Analysis.

In the near-equilibrium regime, one linearizes around a global Maxwellian μ, writing $f = \mu + \sqrt{\mu}h$, and studies the linearized equation:

$$\epsilon\partial_t h + v \cdot \nabla_x h = \frac{1}{\epsilon}\mathcal{L}h, \tag{38}$$

where $\mathcal{L}$ is the linearized Boltzmann operator. Hypercoercivity techniques yield exponential convergence:

$$\|h(t)\|_{L^2} \leq Ce^{-\lambda t}\|h(0)\|_{H^1}, \tag{39}$$

with decay rate $\lambda > 0$ uniform in ϵ.

These techniques also apply to boundary layer analysis, where f must satisfy reflection or absorption boundary conditions. The interplay of collision-driven relaxation and boundary-layer structure is crucial in modeling rarefied gas effects near walls.

Conceptual Significance.

The hydrodynamic limit illustrates a fundamental epistemological transition: from microscopic determinism (kinetic) to macroscopic effective laws (fluid dynamics). Entropy production serves as a mathematical and physical mechanism by which information about individual particle states becomes irrelevant over time. The resulting macroscopic laws capture the emergent, irreversible dynamics.

Hypercoercivity methods enrich this picture by providing a quantitative bridge between scales, offering explicit rates and regularity structures that underlie the smooth passage from stochastic dynamics to deterministic PDEs.

Entropy dissipation ensures compactness and stability in these limits. Furthermore, hypercoercivity provides exponential convergence toward hydrodynamic equilibria in the near-equilibrium regime. These tools are crucial in proving convergence and stability of shock layers and boundary layers in rarefied gases [16,40,41].

6.3. Optimal Transport: Geometry and Functional Inequalities

A striking modern connection arises between kinetic entropy dissipation and *optimal transport theory*. Through the seminal works of Otto, Villani, and others, the space of probability densities $\mathcal{P}_2(\mathbb{R}^d)$, equipped with the 2-Wasserstein distance W_2, inherits a formal Riemannian manifold structure [4,10].

The 2-Wasserstein distance between two probability densities f and g on $\mathbb{R}^d$ is defined by:

$$W_2^2(f,g) := \inf_{\pi \in \Pi(f,g)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \, d\pi(x,y), \tag{40}$$

where $\Pi(f,g)$ is the set of all couplings (joint distributions) with marginals f and g.

In this setting, the Fokker–Planck equation:

$$\partial_t f = \nabla_v \cdot (\nabla_v f + f\nabla_v V), \tag{41}$$

can be viewed as the gradient flow of the free energy functional:

$$\mathcal{F}(f) = \int f \log f \, dv + \int V(v) f(v) \, dv, \tag{42}$$

in the Wasserstein metric space. This geometric insight allows one to understand convergence to equilibrium through the lens of *geodesic convexity*.

A functional $\mathcal{F}$ is said to be λ-convex along Wasserstein geodesics if:

$$\mathcal{F}(f_t) \leq (1-t)\mathcal{F}(f_0) + t\mathcal{F}(f_1) - \frac{\lambda}{2}t(1-t)W_2^2(f_0,f_1), \quad t \in [0,1], \tag{43}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

where f_t is the geodesic interpolation between f_0 and f_1 in $\mathcal{P}_2$.

This convexity implies functional inequalities with deep implications:

- **Logarithmic Sobolev inequality:**

$$\text{Ent}_\gamma(f) \leq \frac{1}{2\lambda} I(f|\gamma), \tag{44}$$

where $\text{Ent}_\gamma(f) = \int f \log \frac{f}{\gamma}$ and $I(f|\gamma) = \int \left|\nabla \log \frac{f}{\gamma}\right|^2 f$ is the Fisher information.

- **Talagrand's inequality:**

$$W_2^2(f, \gamma) \leq 2\text{Ent}_\gamma(f). \tag{45}$$

- **HWI inequality:**

$$\text{Ent}_\gamma(f) \leq W_2(f, \gamma)\sqrt{I(f|\gamma)} - \frac{\lambda}{2} W_2^2(f, \gamma). \tag{46}$$

These inequalities provide quantitative bounds on entropy decay and convergence in W_2. In kinetic theory, they ensure exponential relaxation rates in diffusion equations (e.g., Fokker–Planck), and indirectly control regularity and stability.

Ricci Curvature and the CD(K,N) Condition.

A significant advance was the generalization of Ricci curvature lower bounds to metric measure spaces by Lott–Sturm–Villani [12,22,23]. A space (X, d, m) satisfies the curvature-dimension condition $\text{CD}(K, N)$ if entropy functionals are K-convex along Wasserstein geodesics, mimicking the behavior on smooth Riemannian manifolds with Ricci curvature bounded below by K.

This synthetic notion of curvature underpins the analysis of heat flow and kinetic evolution in non-smooth spaces. It links entropy decay, optimal transport, and geometric analysis into a cohesive framework for understanding convergence in both classical and generalized settings.

Conceptual Analysis.

The optimal transport viewpoint reinterprets dissipation not merely as a loss of information but as a geometric flow in the space of distributions. Entropy becomes a potential, and its decay corresponds to a descent along the steepest path in $\mathcal{P}_2$. This formalism unifies thermodynamic irreversibility, probabilistic dispersion, and geometric curvature into one analytic mechanism.

In kinetic theory, this reveals entropy production as a fundamentally geometric process, tied not only to microscopic collisions but to the intrinsic geometry of mass rearrangement. The Wasserstein metric becomes a powerful tool to quantify and guide such evolution.

These results create a unified geometric and analytic framework to understand stability and long-time behavior in kinetic theory, grounded in convexity and curvature. In particular, the Lott–Sturm–Villani curvature-dimension condition $\text{CD}(K, N)$ formalizes this bridge between kinetic entropy and Ricci curvature [12,22].

6.4. Data Science and Machine Learning

The interplay between kinetic equations and optimal transport theory has recently led to significant advances in data science, particularly in the design and analysis of generative models, diffusion-based algorithms, and sampling methods in high-dimensional probability spaces. These connections draw upon the deep mathematical foundations of entropy dissipation, gradient flows, and variational structures.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

Optimal Transport and Learning.

Let μ and ν be probability measures on $\mathbb{R}^d$. In generative modeling, the objective is often to learn a transport map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $T\sharp\mu = \nu$, i.e., ν is the push-forward of μ through T. The Wasserstein-2 distance:

$$W_2^2(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \|x-y\|^2 \, d\pi(x,y), \tag{47}$$

quantifies the cost of transporting mass from μ to ν. This distance is used to train generative adversarial networks (GANs), such as the Wasserstein GAN [45], by minimizing W_2 between real and generated distributions.

Fokker–Planck and Diffusion Models.

Let $f(t,x)$ denote the density of a random variable evolving under the Langevin dynamics:

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dW_t, \tag{48}$$

where V is a potential function and W_t is a standard Brownian motion. The associated Fokker–Planck equation governing the evolution of f is:

$$\partial_t f = \nabla_x \cdot (\nabla_x f + f\nabla_x V). \tag{49}$$

This PDE is the gradient flow of the free energy functional:

$$\mathcal{F}(f) = \int f \log f + Vf \, dx, \tag{50}$$

in the Wasserstein space $\mathcal{P}_2(\mathbb{R}^d)$ [10,46]. Minimizing $\mathcal{F}$ amounts to sampling from the Gibbs measure $e^{-V(x)}$.

These dynamics underpin score-based generative models and diffusion probabilistic models [47]. The generative process involves solving the reverse-time stochastic differential equation, which corresponds to the adjoint Fokker–Planck evolution.

Entropic Regularization and Sinkhorn Algorithm.

To make optimal transport computationally feasible in high dimensions, entropic regularization is introduced:

$$W_{2,\varepsilon}^2(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \int \|x-y\|^2 \, d\pi(x,y) + \varepsilon \mathrm{KL}(\pi|\mu\otimes\nu), \tag{51}$$

where KL denotes the Kullback–Leibler divergence. The minimizer satisfies a scaling equation that can be efficiently computed using the Sinkhorn algorithm [48]. This approach is widely used in domain adaptation, clustering, and matching tasks.

Kinetic Theory and Sampling Algorithms.

Sampling from complex distributions can be interpreted as evolving particles according to a kinetic equation toward equilibrium. For example, the underdamped Langevin dynamics obey:

$$\begin{cases} dX_t = V_t \, dt, \\ dV_t = -\gamma V_t \, dt - \nabla V(X_t) \, dt + \sqrt{2\gamma} \, dW_t, \end{cases} \tag{52}$$

which corresponds to a kinetic Fokker–Planck equation:

$$\partial_t f + v \cdot \nabla_x f = \nabla_v \cdot (\gamma v f + \nabla_v f + f \nabla_x V). \tag{53}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

Hypercoercivity methods guarantee exponential convergence of f to equilibrium [28,34], making this formulation robust for large-scale Bayesian inference.

Conceptual Insights.

The central unifying idea is that learning and sampling can be interpreted as evolution processes on the space of probability measures. The geometry of this space–particularly when endowed with the Wasserstein metric–guides the formulation of dynamics with provable convergence properties. Entropy and functional inequalities (e.g., logarithmic Sobolev, Talagrand) provide theoretical guarantees for convergence rates and stability.

Hence, tools from kinetic theory and optimal transport are not just analytical devices but also constructive frameworks for algorithm design in machine learning.

6.5. Astrophysics: Long-Time Dynamics of Stellar Systems

In astrophysics, kinetic models play a foundational role in describing the evolution of self-gravitating systems such as galaxies, globular clusters, and dark matter halos. On large spatial and temporal scales, these systems are governed by mean-field interactions through gravity, and their dynamics are well captured by the **Vlasov–Poisson system**:

$$\partial_t f + v \cdot \nabla_x f - \nabla_x \Phi \cdot \nabla_v f = 0, \quad \Delta\Phi = 4\pi G \int f\, dv, \tag{54}$$

where $f = f(t, x, v)$ is the distribution function in phase space and $\Phi = \Phi(t, x)$ is the gravitational potential generated by the mass density $\rho_f(x) = \int f(x, v)\, dv$. The gravitational constant is denoted by G.

This equation system is *collisionless*, i.e., it neglects binary particle interactions in favor of collective field effects. Despite this, the Vlasov–Poisson system preserves several important invariants: mass, momentum, energy, and Casimir functionals such as the Boltzmann entropy:

$$\mathcal{H}(f) = \int f \log f\, dxdv. \tag{55}$$

Although entropy is conserved in the collisionless Vlasov evolution, gravitational systems display phenomena such as *violent relaxation* (as introduced by Lynden-Bell [42]), where systems approach quasi-stationary states on dynamical timescales. These states are thought to approximate maximum entropy configurations under constraints.

Adding Weak Collisions: The Fokker–Planck–Landau Model

To model the long-time collisional relaxation (e.g., in star clusters), one incorporates a weak collisional operator such as the Landau or Fokker–Planck term:

$$\partial_t f + v \cdot \nabla_x f - \nabla_x \Phi \cdot \nabla_v f = \nabla_v \cdot (D(v)\nabla_v f + F(v) f), \tag{56}$$

where $D(v)$ is the diffusion matrix and $F(v)$ is a friction term. This equation conserves mass and energy but now leads to entropy *dissipation*:

$$\frac{d}{dt}\mathcal{H}(f) \le 0. \tag{57}$$

In the long-time regime, entropy dissipation drives convergence toward an equilibrium state that minimizes the free energy functional

$$\mathcal{F}(f) = \int f \log f\, dxdv + \frac{1}{2}\int \Phi_f(x)\rho_f(x)\, dx, \tag{58}$$

 doi:10.20944/preprints202503.2313.v1

under mass and energy constraints. The minimizers are *isothermal spheres* or other steady solutions of the form:

$$\nabla_v f + f\nabla_v\left(\frac{v^2}{2} + \Phi(x)\right) = 0 \quad \Rightarrow \quad f(x,v) \propto \exp\left(-\frac{v^2}{2} - \Phi(x)\right). \tag{59}$$

Macroscopic Limits and Jeans Equations

Integrating the kinetic equation in velocity yields the **Jeans equations**, which are the macroscopic analogues of the Vlasov–Poisson model. For instance, the momentum balance reads:

$$\partial_t(\rho u_i) + \partial_j(\rho u_i u_j + P_{ij}) = -\rho\partial_i\Phi, \tag{60}$$

where P_{ij} is the velocity dispersion tensor. These equations play a central role in dynamical modeling of galaxies and stability analysis.

Entropy methods and energy-Casimir techniques are also used to analyze the stability of steady states, based on Lyapunov functionals that measure deviations from equilibrium [39,43,44].

Conceptual Insights

In contrast to plasma and fluid systems, self-gravitating systems are *non-extensive* and *nonlinear* in a fundamentally different way: the gravitational potential is long-range and the system does not admit local thermodynamic equilibrium. As a result, classical entropy maximization requires reinterpretation. The entropy dissipation mechanisms introduced by weak collisions or coarse-graining serve as effective surrogates, allowing for a statistical description of structure formation.

The connection to kinetic theory emphasizes how gravitational dynamics can be encoded in phase space flows and how geometric methods (e.g., Wasserstein flow for diffusive relaxation) can be generalized to incorporate field-theoretic interactions.

Ultimately, the study of kinetic entropy in astrophysics reveals how statistical behavior, geometric constraints, and field dynamics jointly determine the long-time fate of large-scale cosmic structures.

7. Summary and Future Directions

The interplay between kinetic theory, geometric analysis, and partial differential equations has profoundly enriched the understanding of entropy, irreversibility, and equilibrium. The hypercoercivity method not only resolves degeneracies in phase space but also lays the foundation for quantitative convergence results across diverse applications. Future directions include:

- Nonlinear hypocoercivity in multi-species and reactive systems,
- Entropic regularization in numerical optimal transport,
- Stochastic particle methods preserving entropy dissipation,
- Quantum kinetic theory and entropy in degenerate Fermi gases.

These emerging areas continue to be influenced by the foundational work on entropy dissipation and the geometric structure of kinetic equations.

Applications and Broader Implications

The study of entropy production and stability in kinetic equations has profound implications across multiple scientific domains:

- Plasma physics: The kinetic description of plasmas relies on entropy methods to understand collisional relaxation.
- Astrophysics: The Boltzmann equation models stellar dynamics, including the formation of large-scale structures.
- Fluid dynamics: Understanding kinetic entropy methods has led to new insights into the behavior of compressible and rarefied flows.

The hypercoercivity method has profound applications across multiple fields:

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2025 doi:10.20944/preprints202503.2313.v1

- Plasma physics: Understanding collisional relaxation in magnetized plasmas.
- Astrophysics: Modeling the long-term evolution of stellar systems.
- Fluid dynamics: Establishing decay rates in rarefied and compressible flows.
- Optimal transport: Bridging kinetic theory with Ricci curvature and functional inequalities.

By combining multiple mathematical techniques, hypercoercivity provides a powerful framework for studying stability and convergence in kinetic equations, bridging the gap between microscopic particle dynamics and macroscopic thermodynamics. Moreover, the connection between entropy dissipation and optimal transport theory has opened new directions in geometric analysis, linking the stability of kinetic equations with Ricci curvature and functional inequalities.

8. Conclusion

The theory of collisional kinetic equations–anchored in the Boltzmann equation, entropy production, and the modern framework of hypocoercivity–stands at a unique intersection of mathematical physics, differential geometry, and the epistemology of irreversibility. A central tension in the foundations of statistical mechanics lies in the reconciliation of time-reversible microscopic laws with the observed irreversibility of macroscopic evolution. This tension is exemplified by the Boltzmann equation: derived from Newtonian mechanics through statistical approximations, it leads to the H-theorem, asserting the monotonic increase of entropy.

Epistemologically, this marks a transition from ontological determinism–the belief in complete predictability of systems through trajectories–to a statistical epistemology, where knowledge of the system is encoded in a distribution function $f(t,x,v)$, and physical predictions emerge from averages and aggregate behavior. The H-theorem thus becomes not merely a mathematical identity, but a conceptual bridge: it explains how order emerges from disorder, and how equilibrium arises as a natural statistical attractor in complex systems.

The emergence of hypercoercivity theory, as pioneered by Villani and others, reveals a deeper geometric layer in kinetic equations. While classical coercivity arguments fail due to degeneracy, the commutator structure and Lie algebraic generation of phase space directions allow one to recover control via indirect paths. The presence of hypoellipticity–where regularity propagates through non-commuting vector fields–highlights a crucial insight: geometry is the mediator between locality and globality in PDEs.

More profoundly, the connection to optimal transport theory–and in particular, the Wasserstein geometry of probability measures–provides a conceptual unity between entropy dissipation, transport cost, and curvature. Through the Lott–Sturm–Villani theory, Ricci curvature becomes an analytical tool to measure how entropy behaves along geodesics in space of measures. This reframes classical thermodynamic inequalities (e.g., logarithmic Sobolev, Talagrand) as manifestations of geometric convexity in measure-theoretic settings.

The techniques developed to handle entropy production and hypocoercivity transcend their original context. Whether analyzing collisional plasmas, galactic dynamics, fluid instabilities, or sampling algorithms in machine learning, the same mathematical skeleton recurs: a dissipative operator in velocity, a conservative transport in space, and a structure of indirect coercivity restored through commutators and functional coupling.

This synthesis exemplifies a rare unification in mathematics: tools from microlocal analysis, geometric measure theory, information geometry, and functional inequalities coalesce to form a single conceptual edifice. The result is not only a set of theorems about exponential convergence, but a vision of how disorder evolves, how systems forget their initial data, and how irreversibility becomes embedded in the very fabric of mathematical structure.

From a philosophical perspective, these results illuminate the nature of physical law. The move from deterministic particle trajectories to probabilistic evolution via PDEs represents an epistemological concession: we do not claim to know individual microstates, but instead describe their collective effect with statistical certitude. Entropy, in this light, becomes not merely a measure of disorder,

 doi:10.20944/preprints202503.2313.v1

but a quantifier of epistemic irreducibility: it formalizes the limits of knowledge about individual constituents.

Moreover, the existence of an entropy functional whose dissipation governs convergence to equilibrium exemplifies a principle of directionality in time–an emergent arrow not present in the microscopic dynamics. This offers a mathematical framework for grounding the thermodynamic time asymmetry, one of the most enduring puzzles in the philosophy of physics.

The study of entropy dissipation and convergence in kinetic theory–especially via hypercoercivity–reveals a deep and multifaceted structure at the intersection of geometry, analysis, and physics. Far from being a technical tool, entropy becomes a conceptual lens through which the transition from micro to macro, from determinism to irreversibility, from geometry to evolution, is both mathematically expressible and epistemologically coherent.

In recent years, the mathematical techniques developed in kinetic theory and optimal transport have found profound applications beyond traditional physical systems. In particular, the geometry of the space of probability measures, functional inequalities such as logarithmic Sobolev and Talagrand-type inequalities, and the analysis of gradient flows in Wasserstein space have become foundational in modern data science. Applications include sampling algorithms, variational inference, and generative models such as Wasserstein GANs and score-based diffusion models. These developments mirror, at a computational and conceptual level, the dynamics studied in kinetic theory and reveal a deep structural analogy between thermodynamic relaxation and statistical learning.

The mathematical structures presented here not only affirm Boltzmann's vision but extend it into new realms of geometric analysis, with applications that continue to expand into quantum theory, data science, and the very foundations of thermodynamic law. This paper has sought to explore the mathematical mechanisms underlying entropy dissipation, regularization, and convergence to equilibrium, and to reveal how these mechanisms encode deep structural properties of both microscopic dynamics and macroscopic thermodynamics

Author Contributions: All authors contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding'.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

OT	Optimal Transport
GAN	Generative Adversarial Networks

References

1. G. Monge, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l'Académie Royale des Sciences, Paris, 1781.
2. L. V. Kantorovich, *On the translocation of masses*, Dokl. Akad. Nauk. USSR, 37:227–229, 1942.
3. C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics, Vol. 58, American Mathematical Society, 2003.
4. C. Villani, *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften, Vol. 338, Springer, 2009.
5. R. Dobrushin, *Prescribing a system of random variables by conditional distributions*, Theory of Probability and its Applications, 15(3):458–486, 1970.
6. H. Tanaka, *Probabilistic treatment of the Boltzmann equation of Maxwellian molecules*, Z. Wahrscheinlichkeitstheorie verw. Gebiete, 46:67–105, 1978.
7. J. N. Mather, *Action minimizing invariant measures for positive definite Lagrangian systems*, Math. Z., 207:169–207, 1991.

 doi:10.20944/preprints202503.2313.v1

8. Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Communications on Pure and Applied Mathematics, 44(4):375–417, 1991.
9. M. J. P. Cullen and M. R. D. Purser, *An extended Lagrangian theory of semi-geostrophic frontogenesis*, J. Atmos. Sci., 41:1477–1497, 1984.
10. F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Communications in Partial Differential Equations, 26:101–174, 2001.
11. L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Birkhäuser, 2008.
12. J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Annals of Mathematics, 169(3):903–991, 2009.
13. M. Arjovsky, S. Chintala, and L. Bottou, *Wasserstein GAN*, Proc. of ICML, 2017.
14. G. Peyré and M. Cuturi, *Computational Optimal Transport*, Foundations and Trends in Machine Learning, 11(5–6):355–607, 2019.
15. V. M. Panaretos and Y. Zemel, *An Invitation to Statistics in Wasserstein Space*, Springer, 2020.
16. Golse, F. (2005). "The Boltzmann Equation and Its Hydrodynamic Limits." In Handbook of Mathematical Fluid Dynamics, Vol. 3, edited by S. Friedlander and D. Serre, 121–205. Elsevier.
17. M. Cuturi, *Sinkhorn Distances: Lightspeed Computation of Optimal Transport*, Advances in Neural Information Processing Systems (NeurIPS), 2013.
18. L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, *Scaling Algorithms for Unbalanced Optimal Transport Problems*, Mathematics of Computation, 87(314):2563–2609, 2018.
19. J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem*, Numerische Mathematik, 84:375–393, 2000.
20. M. Agueh and G. Carlier, *Barycenters in the Wasserstein space*, SIAM Journal on Mathematical Analysis, 43(2):904–924, 2011.
21. R. McCann, *A convexity principle for interacting gases*, Adv. Math., 128(1):153–179, 1997.
22. K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta Math., 196(1):65–131, 2006.
23. K.-T. Sturm, *On the geometry of metric measure spaces. II*, Acta Math., 196(1):133–177, 2006.
24. M. Talagrand, *Transportation cost for Gaussian and other product measures*, Geom. Funct. Anal., 6(3):587–600, 1996.
25. F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal., 173(2):361–400, 2000.
26. C. Villani, *A review of mathematical topics in collisional kinetic theory*, in Handbook of Mathematical Fluid Dynamics, vol. 1, 2002.
27. C. Cercignani, *The Boltzmann Equation and Its Applications*, Springer, 1994.
28. C. Villani, *Hypocoercivity*, Memoirs of the AMS, 202(950), 2009.
29. F. Golse, *The Boltzmann equation and its hydrodynamic limits*, in Evolutionary Equations, Handbook of Differential Equations, Elsevier, 2013.
30. C. Mouhot and C. Villani, *On Landau damping*, Acta Mathematica, 207(1):29–201, 2011.
31. C. Cercignani, *H-theorem and trend to equilibrium in the kinetic theory of gases*, Arch. Mech., 34(3):231–241, 1982.
32. C. Villani and G. Toscani, *Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the Boltzmann equation*, Communications in Mathematical Physics, 203(3):667–706, 1999.
33. L. Hörmander, *Hypoelliptic second order differential equations*, Acta Mathematica, 119(1):147–171, 1967.
34. F. Hérau, *Hypocoercivity and exponential time decay for the linear inhomogeneous relaxation Boltzmann equation*, Asymptotic Analysis 46.3-4 (2006): 349–359.
35. J. Dolbeault, C. Mouhot, and C. Schmeiser, *Hypocoercivity for kinetic equations with linear relaxation terms*, Comptes Rendus Mathématique, 347(9–10):511–516, 2009.
36. C. Mouhot, *Rate of convergence to equilibrium for the spatially homogeneous Boltzmann equation with hard potentials*, Communications in Mathematical Physics, 261(3):629–672, 2006.
37. R. Glassey, *The Cauchy Problem in Kinetic Theory*, SIAM, 1996.
38. P. Degond and S. Mas-Gallic, *The weighted particle method for convection-diffusion equations. Part 1: The case of an isotropic viscosity*, Mathematics of Computation, 53(188):485–507, 1989.
39. J. Binney and S. Tremaine, *Galactic Dynamics*, Princeton University Press, 2008.
40. C. Bardos, F. Golse, and D. Levermore, *Fluid dynamic limits of kinetic equations I: Formal derivations*, Journal of Statistical Physics, 63(1–2):323–344, 1991.
41. L. Saint-Raymond, *Hydrodynamic Limits of the Boltzmann Equation*, Springer, 2009.

 doi:10.20944/preprints202503.2313.v1

42. D. Lynden-Bell, Statistical mechanics of violent relaxation in stellar systems, *Monthly Notices of the Royal Astronomical Society*, 136(1):101–121, 1967.
43. G. Rein and A.D. Rendall, Smooth static solutions of the Vlasov–Einstein system, *Annales de l'I.H.P. Physique théorique*, 59(4):383–397, 1993.
44. Y. Guo and G. Rein, Isotropic steady states in galactic dynamics, *Communications in Mathematical Physics*, 219(3):607–629, 2001.
45. Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. *International Conference on Machine Learning (ICML)*.
46. Jordan, R., Kinderlehrer, D., & Otto, F. (1998). The variational formulation of the Fokker–Planck equation. *SIAM Journal on Mathematical Analysis*, 29(1), 1–17.
47. Song, Y., & Ermon, S. (2021). Score-Based Generative Modeling through Stochastic Differential Equations. *International Conference on Learning Representations (ICLR)*.
48. Cuturi, M. (2013). Sinkhorn Distances: Lightspeed Computation of Optimal Transport. *Advances in Neural Information Processing Systems (NeurIPS)*, 26.