Journal of Statistical Physics, Vol. 94, Nos. 3/4, 1999

Probability Metrics and Uniqueness of the Solution to the Boltzmann Equation for a Maxwell Gas

G. Toscani[1] and C. Villani[2]

Received December 23, 1997; final September 28, 1998

We consider a metric for probability densities with finite variance on $\mathbb{R}^d$, and compare it with other metrics. We use it for several applications both in probability and in kinetic theory. The main application in kinetic theory is a uniqueness result for the solution of the spatially homogeneous Boltzmann equation for a gas of true Maxwell molecules.

KEY WORDS: Spatially homogeneous Boltzmann equation; probability metrics; Maxwellian molecules.

1. INTRODUCTION

Denote by $P_s(\mathbb{R}^d)$, $s > 0$, the class of all probability distributions F on $\mathbb{R}^d$, $d \geqslant 1$, such that

$$\int_{\mathbb{R}^d} |v|^s \, dF(v) < \infty$$

We introduce a metric on $P_s(\mathbb{R}^d)$ by

$$d_s(F, G) = \sup_{\xi \in \mathbb{R}^d} \frac{|\hat{f}(\xi) - \hat{g}(\xi)|}{|\xi|^s} \tag{1}$$

where $\hat{f}$ is the Fourier transform of F,

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} e^{-i\xi \cdot v} \, dF(v)$$

[1] Department of Mathematics, University of Pavia, 27100 Pavia, Italy.
[2] Ecole Normale Supérieure, DMI, 75230 Paris Cedex 05, France.

0022-4715/99/0200-0619$16.00/0

Let us write $s = m + \alpha$, where m is an integer and $0 \leqslant \alpha < 1$. In order that $d_s(F, G)$ be finite, it suffices that F and G have the same moments up to order m.

The norm (1) has been introduced in ref. 6 to investigate the trend to equilibrium of the solutions to the Boltzmann equation for Maxwell molecules. There, the case $s = 2 + \alpha$, $\alpha > 0$, was considered. Further applications of d_s, with $s = 4$, were studied in ref. 3, while the cases $s = 2$ and $s = 2 + \alpha$, $\alpha > 0$, have been considered in ref. 4 in connection with the so-called Mc Kean graphs.[8]

In this paper, we shall be interested with the case $s = 2$. To understand why this case separates in a natural way from the other ones, let us briefly introduce and discuss other well-known metrics on $P_s(\mathbb{R}^d)$.

Let F, G in $P_s(\mathbb{R}^d)$, and let $\Pi(F, G)$ be the set of all probability distributions L in $P_s(\mathbb{R}^d \times \mathbb{R}^d)$ having F and G as marginal distributions. Let

$$T_s(F, G) = \inf_{L \in \Pi(F, G)} \int |v - w|^s \, dL(v, w) \tag{2}$$

Then $\tau_s = T_s^{1/s}$ metrizes the weak-* topology TW_* on $P_s(\mathbb{R}^d)$. We note that T_1 is the Kantorovich–Vasershtein distance of F and G.[7, 18] For a detailed discussion, and application of these distances to statistics and information theory, see Vajda.[17] See also ref. 10 for a recent application to kinetic theory.

The case $s = 2$ was introduced and studied independently by Tanaka[15] who, in the one-dimensional case $d = 1$, applied T_2 to the study of Kac's equation. Subsequently, the properties of T_2 were studied in the multidimensional case by Murata and Tanaka.[9] Applications to the kinetic theory of rarefied gases were finally studied by Tanaka in ref. 16; several of these applications were given a simplified proof in ref. 12.

The importance of Tanaka's distance τ_2 mainly relies on its convexity and superadditivity with respect to rescaled convolutions. We recall this property, that is at the basis of most of the applications of T_2.

Let $\{X_0, Y_0\}$, $\{X_1, Y_1\}$ be two independent pairs of random variables, and let F_i (resp. G_i) be the probability distribution of X_i (resp. Y_i), $i = 0, 1$. For $0 < \lambda < 1$, let F_λ (resp. G_λ) be the probability distribution of $\sqrt{\lambda}\, X_0 + \sqrt{1-\lambda}\, X_1$ (resp. $\sqrt{\lambda}\, Y_0 + \sqrt{1-\lambda}\, Y_1$), i.e.

$$F_\lambda = \frac{1}{\lambda^{d/2}} F_0\left(\frac{\cdot}{\sqrt{\lambda}}\right) * \frac{1}{(1-\lambda)^{d/2}} F_1\left(\frac{\cdot}{\sqrt{1-\lambda}}\right) \tag{3}$$

Then,

$$T_2(F_\lambda, G_\lambda) \leqslant \lambda T_2(F_0, G_0) + (1-\lambda)\, T_2(F_1, G_1) \tag{4}$$

Superadditivity is also known for convex functionals (relative entropies), like Boltzmann's relative entropy

$$H(f \mid M^f) = \int_{\mathbb{R}^d} f(v) \log \frac{f(v)}{M^f(v)}\, dv \tag{5}$$

where f is a probability density and M^f is the Gaussian density with the same mean vector and variance as those of f. This means that the property (4) holds with T_2 replaced by H and $\{G_0, G_1\}$ replaced by $\{M^{f_0}, M^{f_1}\}$. This is a consequence of Shannon's entropy power inequality (Cf. refs. 13 and 1). The same property holds for the relative Fisher information,

$$I(f \mid M^f) = \int_{\mathbb{R}^d} |\nabla \log f(v) - \nabla \log M^f(v)|^2\, f(v)\, dv \tag{6}$$

(see again refs. 13 and 1). As discussed by Csiszar,[5] by means of the relative entropy H, one can define the so-called H-neighbourhoods. Even if those do not define a topological space, in the usual sense, their topological structure is finer than the metric topology defined by the total variation distance,

$$v(f, g) = \int |f(v) - g(v)|\, dv$$

in the sense that

$$v(f, g) \leqslant \sqrt{2H(f \mid g)} \tag{7}$$

which is the so-called Csiszar–Kullback inequality.

It turns out that these properties of superadditivity and convexity also hold for d_2, the proofs being in fact much more simple. As an illustration of the interest of these properties, we shall give a version of the central limit theorem and a very simple proof of Kac's theorem.

We shall also apply d_2 to the study of the Boltzmann equation with Maxwellian molecules,

$$\begin{aligned}\frac{\partial f}{\partial t}(t, v) &= Q(f, f)(t, v)\\ &= \int_{\mathbb{R}^3 \times S^2} \sigma\left(\frac{u \cdot n}{|u|}\right) [f(v')\, f(w') - f(v)\, f(w)]\, dw\, dn\end{aligned} \tag{8}$$

where $u = v - w$ is the relative velocity of colliding particles with velocity v and w, and

$$v' = \frac{v+w}{2} + \frac{|u|}{2} n, \qquad w' = \frac{v+w}{2} - \frac{|u|}{2} n$$

are the postcollisional velocities. $\sigma(v)$ is a nonnegative function which for true Maxwell molecules has a nonintegrable singularity of the form $(1-v)^{-5/4}$ as $v \to 1$. Usually, one truncates σ in some way, so that it become integrable (cut-off assumption).

We shall prove that d_2 shares a remarkable property with Tanaka's distance: it is nonexpanding with time along trajectories of the Boltzmann equation; that is, if f and g are two such solutions,

$$d_2(f(t), g(t)) \leqslant d_2(f(0), g(0)) \tag{9}$$

This holds even if σ is singular. As an immediate corollary, we obtain that the solution to the Cauchy problem for the Boltzmann equation is unique. Up to our knowledge, this is so far the only uniqueness result available for long-range interactions.

The organization of the paper is as follows. First, in Section 2, we investigate the connections between several distances on $P_2(\mathbb{R}^d)$, including d_2 and τ_2. In Section 3, we establish the basic properties of superadditivity for d_2 and give applications. Then, in Section 4, we apply this metric to the study of the Boltzmann equation.

2. METRICS ON $P_2(\mathbb{R}^d)$

In order that d_2 be well-defined, we need to restrict it to some space of probability densities with the same mean vector. For simplicity, we shall restrict to probability measures with zero mean vector, and we shall work on

$$D_\sigma = \left\{ F \in P_2(\mathbb{R}^d);\ \int v_i\, dF(v) = 0,\ \int |v|^2\, dF(v) = d \cdot \sigma \right\} \tag{10}$$

where σ is some positive real number. We begin with two elementary lemmas.

Lemma 1. Let (F_n) in D_σ, $F_n \rightharpoonup F$ in $P_0(\mathbb{R}^d)$. Then

$$F \in D_\sigma \Leftrightarrow \lim_{K \to \infty} \sup_{n \to \infty} \int |v|^2\, 1_{|v| \geqslant K}\, dF_n(v) = 0 \tag{11}$$

Proof. It is clear that if $\lim_{K \to \infty} \sup_{n \to \infty} \int |v|^2 1_{|v| \geqslant K} dF_n(v) = 0$, then $F \in D_\sigma$. On the other hand, if this condition is not satisfied, then there exists $\varepsilon > 0$ and $(K_n) \to \infty$ with $\int |v|^2 1_{|v| \leqslant K_n} dF_n(v) \leqslant d \cdot \sigma - \varepsilon$. Since $|v|^2 1_{|v| \leqslant K_n} \rightharpoonup |v|^2$, this implies that $\int |v|^2 dF(v) \leqslant d\sigma - \varepsilon$, hence $F \notin D_\sigma$. ∎

Lemma 2. Let (F_n) and F in D_σ, such that $F_n \rightharpoonup F$. Then, there exists a nonnegative function ϕ such that $\phi(r)/r \to \infty$ as $r \to \infty$, which can be chosen smooth and convex, and a constant $M > 0$, such that

$$\sup_n \int \phi(|v|^2) \, dF_n(v) \leqslant M \tag{12}$$

Proof. Using Lemma 1, copy the construction that was done in ref. 6 for one single function: there exists k_1 such that $\sup_n \int_{|v| \geqslant k_1} |v|^2 \, dF_n(v) \leqslant 1/2$; then there exists $k_2 \geqslant k_1$ such that $\sup_n \int_{|v| \geqslant k_2} |v|^2 \, dF_n(v) \leqslant 1/4$; and so on... for all $p \geqslant 2$, there exists $k_p \geqslant k_{p-1}$ such that

$$\sup_n \int_{|v| \geqslant k_p} |v|^2 \, dF_n(v) \leqslant \frac{1}{2^p}$$

Then choose $\phi(|v|^2) = p \, |v|^2$ if $|v| \in [k_p, k_{p+1})$; smooth this function and slow its growth if necessary, as in ref. 6. ∎

In the sequel of the paper, for $M > 0$, we shall denote by

$$P^M_{2+\alpha}(\mathbb{R}^d) = \left\{ F \in D_\sigma; \int |v|^{2+\alpha} \, dF(v) \leqslant M \right\} \tag{13}$$

$$P^M_\phi(\mathbb{R}^d) = \left\{ F \in D_\sigma; \int \phi(|v|^2) \, dF(v) \leqslant M \right\} \tag{14}$$

for ϕ nonnegative, $\phi(r)/r \to \infty$ as $r \to \infty$. Lemma 2 enables us to restrict to spaces P^M_ϕ in all the cases when one is interested with weak convergence in D_σ.

In addition to the metrics d_2 and $\tau = \tau_2$ which were introduced in the last section, we consider

- *Prokhorov's distance* $\rho(F, G)$: for $\delta \geqslant 0$ and $U \subset \mathbb{R}^d$, we define

$$U^\delta = \{ v \in \mathbb{R}^d; d(v, U) < \delta \}, \qquad U^{\delta]} = \{ v \in \mathbb{R}^d; d(v, U) \leqslant \delta \}$$

where $d(v, U) = \inf\{ \|v - w\|, w \in U \}$. Let

$$\sigma(F, G) = \inf\{ \varepsilon > 0 / F(A) \leqslant G(A^\varepsilon) + \varepsilon \text{ for all closed } A \subset \mathbb{R}^d \}$$

we set

$$\rho(F, G) = \max\{\sigma(F, G), \sigma(G, F)\} \tag{15}$$

• *the* $(C^m)^*$ *distance* $\|F-G\|_m^*$: for $m \geqslant 1$, let $C^m(\mathbb{R}^d)$ be the set of m-times continuously differentiable functions, endowed with its natural norm $\|\cdot\|_m$. Then let

$$\|F-G\|_m^* = \sup\left\{\left|\int \varphi\, dF(v)\right|, \varphi \in C^m, \|\varphi\|_m \leqslant 1\right\} \tag{16}$$

Theorem 1. Let (F_n) in D_σ and F in $P_2(\mathbb{R}^d)$. Then, the following statements are equivalent.

(i) $F_n \rightharpoonup F$ and $F \in D_\sigma$;

(ii) $\tau(F_n, F) \to 0$;

(iii) $\rho(F_n, F) \to 0$ and $F \in D_\sigma$;

(iv) For any $m \geqslant 1$, $\|F_n - F\|_m^* \to 0$, and $F \in D_\sigma$;

(v) $d_2(F_n, F) \to 0$.

Proof. Here we shall only show (i) $\Leftrightarrow$ (v). First, if $d_2(F_n, F) \to 0$, then obviously, for all $\xi \in \mathbb{R}^d \setminus \{0\}$, $\hat{f}_n(\xi) \to \hat{f}(\xi)$; on the other hand, $\hat{f}_n(0) = \hat{f}(0) = 1$. This entails that $F_n \rightharpoonup F$; it remains to prove that $F \in D_\sigma$. But for fixed ξ, $|\xi| = 1$, $t \geqslant 0$, since the first derivatives of $\hat{f}_n$ and $\hat{f}$ coincide at the origin, and since their Fourier transforms are twice continuously differentiable,

$$|D^2(\hat{f}_n - \hat{f})(0) \cdot (\xi, \xi)| = \left|\lim_{t \to 0} \frac{\hat{f}_n(t\xi) - \hat{f}(t\xi)}{t^2}\right| \leqslant d_2(F_n, F) \to 0$$

Conversely, suppose that $F_n \rightharpoonup F \in D_\sigma$. By Lemma 2, there exists ϕ and M such that $F_n \in P_\phi^M$. By Lemma 3.1 of ref. 6, all $D^2\hat{f}_n$ have a uniform modulus of continuity. In particular, for $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\forall n\ |\xi| \leqslant \delta \Rightarrow |D^2\hat{f}_n(\xi) - D^2\hat{f}_n(0)| \leqslant \varepsilon$$

Since

$$\begin{aligned}
\frac{\hat{f}_n(\xi) - \hat{f}(\xi)}{|\xi|^2} = &\int_0^1 [D^2\hat{f}_n(t\xi) - D^2\hat{f}_n(0)] \cdot \left(\frac{\xi}{|\xi|}, \frac{\xi}{|\xi|}\right)(1-t)\, dt \\
&+ \frac{1}{2}[D^2\hat{f}_n(0) - D^2\hat{f}(0)]\left(\frac{\xi}{|\xi|}, \frac{\xi}{|\xi|}\right) \\
&- \int_0^1 [D^2\hat{f}(t\xi) - D^2\hat{f}(0)] \cdot \left(\frac{\xi}{|\xi|}, \frac{\xi}{|\xi|}\right)(1-t)\, dt
\end{aligned}$$

we obtain

$$\sup_{|\xi| \leqslant \delta} \frac{|\hat{f}_n(\xi) - \hat{f}(\xi)|}{|\xi|^2} \leqslant \varepsilon$$

On the other hand, clearly, there exists $D > 0$ such that

$$|\xi| \geqslant D \Rightarrow \frac{|\hat{f}_n(\xi) - \hat{f}(\xi)|}{|\xi|^2} \leqslant \varepsilon$$

Thus $d_2(F_n, F) \leqslant \max\{\varepsilon, \sup_{\delta \leqslant |\xi| \leqslant D} (|\hat{f}_n(\xi) - \hat{f}(\xi)|)/|\xi|^2\} \leqslant \max\{\varepsilon, \sup_{\delta \leqslant |\xi| \leqslant D} 1/\delta^2 \, |\hat{f}_n(\xi) - \hat{f}(\xi)|\}$. Now, since $F_n \rightharpoonup F$ we have

$$\forall \xi, \hat{f}_n(\xi) \to \hat{f}(\xi)$$

and since $|D^2 \hat{f}_n(\xi)| \leqslant d\sigma$ and $D\hat{f}_n(0) = 0$, $(\hat{f}_n)$ is uniformly equicontinuous on the compact set $\{\delta \leqslant |\xi| \leqslant D\}$. By Ascoli's theorem, this entails that $\sup_{\delta \leqslant |\xi| \leqslant D} |\hat{f}_n(\xi) - \hat{f}(\xi)|$ goes to 0, and thus there exists $n_0 \geqslant 0$ such that for $n \geqslant n_0$, $d_2(F_n, F)) \leqslant \varepsilon$. ∎

Next, we would like to compare more precisely these different metrics. We shall say that two metrics m_1 and m_2 define the same weak-* uniformity on a set $S \subset P_2(\mathbb{R}^d)$ if for all $\varepsilon > 0$ there exists $\eta > 0$ such that for all $F, G \in S$,

$$m_1(F, G) \leqslant \eta \Rightarrow m_2(F, G) \leqslant \varepsilon,$$
$$m_2(F, G) \leqslant \eta \Rightarrow m_1(F, G) \leqslant \varepsilon.$$

Theorem 2. Let $M > 0$ and ϕ be fixed. Then, for any $m \geqslant 1$, τ, ρ, $\|\cdot\|_m^*$ and d_2 define the same weak-* uniformity in P_ϕ^M.

In order to simplify the proof, we shall prove this theorem only on $P_{2+\alpha}^M$, where the bounds are explicit. In all the sequel $\alpha > 0$, $M > 0$ and $m \geqslant 1$ will be fixed. The main part of the proof has already been performed in ref. 6, therefore we shall only "fill the gaps." We split the proof in three steps.

First Step. τ and ρ define the same weak-* uniformity on $P_{2+\alpha}^M$. More precisely,

$$\begin{aligned} &\text{(a)} \quad \tau(F, G)^2 \leqslant (2M + 8)\, \rho(F, G)^{\alpha/(\alpha+2)} + 4\rho(F, G)^2 \\ &\text{(b)} \quad \rho(F, G) \leqslant \tau(F, G)^{3/2} \end{aligned} \tag{17}$$

Proof. Part (a) is Theorem 5.2 of ref. 6. As for part (b), let $\beta > 0$, then there exists $L^* \in \Pi(F, G)$ such that

$$T_2(F, G) = \int |v - w|^2 \, dL^*(v, w) \geqslant \beta^2 L^*(|v - w| \geqslant \beta)$$

Chosing $\beta = T_2(F, G)^{1/3}$, we see that

$$L^*(|v - w| \geqslant \beta) \leqslant \beta$$

By Theorem 5.1 of ref. 6, this implies that $F(A) \leqslant G(A^{\delta]}) + \beta$ for all closed $A \subset \mathbb{R}^d$. Thus, $\rho(F, G) \leqslant \beta = \tau(F, G)^{3/2}$. ∎

Second Step. For all $m \geqslant 1$, ρ and $\|\cdot\|_m^*$ define the same weak-* uniformity on $P_0(\mathbb{R}^d)$.

This is done by putting together Lemma 5.3 and Corollary 5.5 in ref. 6.

Third Step. For all $m \geqslant 1$, $\|\cdot\|_m^*$ and d_2 define the same weak-* uniformity on $P_{2+\alpha}^M$. More precisely,

$$\begin{aligned} &\text{(a)} \quad \|F - G\|_{d+3}^* \leqslant C_d \sigma^{(d+1)/(d+3)} d_2(F, G)^{2/(d+3)} \\ &\text{(b)} \quad d_2(F, G) \leqslant C_d M^{2/(2+\alpha)} (\|F - G\|_1^*)^{\alpha/(2+\alpha)} \end{aligned} \tag{18}$$

Proof. (a) is an easy adaptation of Lemma 5.7 in ref. 6. Let $\varphi \in C^{d+3}(\mathbb{R}^d)$, $\|\varphi\|_{d+3} \leqslant 1$; let $R \geqslant 1$. Let χ_R be a smooth function such that $0 \leqslant \chi_R \leqslant 1$, $\chi_R = 1$ for $|v| \leqslant R$, $\chi_R(v) = 0$ for $|v| \geqslant R + 1$. We estimate first the tails of the distributions.

$$\left| \int (1 - \chi_R) \, \varphi d(F - G) \right| \leqslant \int_{|v| \geqslant R} dF + \int_{|v| \geqslant R} dG \leqslant \frac{2\sigma}{R^2}$$

Then, by Parseval,

$$\begin{aligned} &\left| \int \chi_R \varphi d(F - G) \right| \\ &\quad = \left| \int \widehat{\varphi \chi_R}(\xi) [\hat{f}(\xi) - \hat{g}(\xi)] \, d\xi \right| \\ &\quad \leqslant \sup_{\mathbb{R}^d} \frac{|\hat{f}(\xi) - \hat{g}(\xi)|}{|\xi|^2} \sup_{\mathbb{R}^d} [\widehat{\chi_R \varphi}(\xi)(1 + |\xi|^{d+3})] \int \frac{|\xi|^2}{1 + |\xi|^{d+3}} \, d\xi \end{aligned}$$

Using the classical inequality

$$\sup_{\xi} \{(1+|\xi|^m)\, |\widehat{\chi_R \varphi}(\xi)|\} \leqslant C_d \sup_{v} \{(1+|v|)^{d+1}\, |D^m(\chi_R \varphi)(v)|\} \leqslant C_d (1+R)^{d+1}$$

we conclude that

$$\left| \int \varphi d(F-G) \right| \leqslant \frac{2\sigma}{R^2} + C_d R^{d+1} d_2(F, G)$$

Optimizing over R, we get the desired result.

As for (b): clearly,

$$\frac{|\hat{f}(\xi) - \hat{g}(\xi)|}{|\xi|^2} \leqslant \left| \int \frac{\cos(\xi \cdot v)}{|\xi|^2}\, d(F-G) \right| + \left| \int \frac{\sin(\xi \cdot v)}{|\xi|^2}\, d(F-G) \right|$$

Let us estimate for instance the first term in the right-hand side. First we note that

$$\int \frac{\cos(\xi \cdot v)}{|\xi|^2}\, d(F-G) = \int \frac{\cos(\xi \cdot v) - 1}{|\xi|^2}\, d(F-G)$$

For fixed ξ,

$$\Phi_\xi(v) = \frac{\cos(\xi \cdot v) - 1}{|\xi|^2}$$

is a C^2 function, vanishing at the origin, as well as its first-order derivatives, and elementary estimates show that $|\Phi'_\xi(v)| \leqslant |v|$, $\Phi_\xi(v) \leqslant |v|^2/2$, hence $\|\Phi_\xi \chi_R\|_1 \leqslant CR^2$, whence by definition

$$\frac{1}{CR^2} \left| \int \Phi_\xi \chi_R d(F-G) \right| \leqslant \|F-G\|_1^*$$

On the other hand,

$$\left| \int (1-\chi_R)\, \Phi_\xi d(F-G) \right| \leqslant \frac{2M}{R^\alpha}$$

Optimizing over R, we get the result. ∎

3. SUPERADDITIVITY

Let $\{X_0, Y_0\}$ and $\{X_1, Y_1\}$ be two independent pairs of random variables with zero mean vector. Let F_i (resp. G_i) denote the law of X_i (resp. Y_i). For $0<\lambda<1$, the law of

$$X_\lambda = \sqrt{\lambda}\, X_0 + \sqrt{1-\lambda}\, X_1$$

is

$$F_\lambda = \frac{1}{\lambda^{d/2}} F_0\left(\frac{\cdot}{\sqrt{\lambda}}\right) * \frac{1}{(1-\lambda)^{d/2}} F_1\left(\frac{\cdot}{\sqrt{1-\lambda}}\right) \tag{19}$$

Theorem 3. d_2 is superadditive with respect to the rescaled convolution, i.e.

$$d_2(F_\lambda, G_\lambda) \leqslant \lambda d_2(F_0, G_0) + (1-\lambda)\, d_2(F_1, G_1) \tag{20}$$

Corollary. The rescaled convolution is continuous with respect to d_2: with obvious notations, if $d_2(F_0^n, G_0) \to 0$ and $d_2(F_1^n, G_1) \to 0$, then $d_2(F_\lambda^n, G_\lambda) \to 0$.

Proof. We denote by $\hat{f}_i$ and $\hat{g}_i$ the Fourier transforms of F_i and G_i. Since

$$\hat{f}_\lambda(\xi) = \hat{f}_0(\sqrt{\lambda}\,\xi)\, \hat{f}_1(\sqrt{1-\lambda}\,\xi)$$

and an analogous formula holds for $\hat{g}_\lambda$, we have

$$\begin{aligned} d_2(F_\lambda, G_\lambda) &= \sup_\xi \frac{|\hat{f}_0(\sqrt{\lambda}\,\xi)\, \hat{f}_1(\sqrt{1-\lambda}\,\xi) - \hat{g}_0(\sqrt{\lambda}\,\xi)\, \hat{g}_1(\sqrt{1-\lambda}\,\xi)|}{|\xi|^2} \\ &= \sup_\xi \left\{ (1-\lambda)\, |\hat{f}_0(\sqrt{\lambda}\,\xi)|\, \frac{|\hat{f}_1(\sqrt{1-\lambda}\,\xi) - \hat{g}_1(\sqrt{1-\lambda}\,\xi)|}{(1-\lambda)\, |\xi|^2} \right. \\ &\quad \left. + \lambda\, |\hat{g}_1(\sqrt{1-\lambda}\,\xi)|\, \frac{|\hat{f}_0(\sqrt{\lambda}\,\xi) - \hat{g}_0(\sqrt{\lambda}\,\xi)|}{\lambda\, |\xi|^2} \right\} \end{aligned}$$

Since $\|\hat{f}_0\|_\infty \leqslant 1$ and $\|\hat{g}_1\|_\infty \leqslant 1$, this expression can be bounded by

$$(1-\lambda)\, d_2(F_1, G_1) + \lambda d_2(F_0, G_0) \qquad \blacksquare$$

Remark. This property is sufficient to imply that d_2 is nonexpandive along solutions of the Boltzmann equation if $d=2$ (Cf. ref. 2). But we shall show in the next section that this restriction can be dispended with.

As an application, for X a random variable with law F, let us consider the functional

$$J(X) = J(F) = d_2(F, M^F) = \sup_{\xi} \frac{|\hat{f} - \widehat{M^F}|}{|\xi|^2} \tag{21}$$

where this time M^F is the Gaussian distribution with the same mean vector and *covariance matrix* as F. The same proof as before shows the

Theorem 4. For any two independent random variables X_0 and X_1, and $0 < \lambda < 1$,

$$J(\sqrt{\lambda}\, X_0 + \sqrt{1-\lambda}\, X_1) \leqslant \lambda J(X_0) + (1-\lambda)\, J(X_1) \tag{22}$$

with equality if and only if X_0 and X_1 are gaussian variables with the same mean vector and covariance matrix.

Proof. Suppose that there is equality in the proof of Theorem 3, with G_0 and G_1 replaced by a centered gaussian probability law (one can always reduce to this case). Suppose that $F_0 \neq M^{F_0}$. Let us denote by g_0 and g_1 the densities of M^{F_0} and M^{F_1}. Let (ξ_n) such that

$$\frac{|\hat{f}_0(\sqrt{\lambda}\, \xi_n)\, \hat{f}_1(\sqrt{1-\lambda}\, \xi_n) - \hat{g}_0(\sqrt{\lambda}\, \xi_n)\, \hat{g}_1(\sqrt{1-\lambda}\, \xi_n)|}{|\xi_n|^2}$$

$$\xrightarrow{n \to \infty} J(\sqrt{\lambda}\, X_0 + \sqrt{1-\lambda}\, X_1)$$

the left-hand side is bounded by

$$\left\{ (1-\lambda)\, |\hat{f}_0(\sqrt{\lambda}\, \xi_n)|\, \frac{|\hat{f}_1(\sqrt{1-\lambda}\, \xi_n) - \hat{g}_1(\sqrt{1-\lambda}\, \xi_n)|}{(1-\lambda)\, |\xi_n|^2} \right.$$

$$\left. + \lambda\, |\hat{g}_1(\sqrt{1-\lambda}\, \xi_n)|\, \frac{|\hat{f}_0(\sqrt{\lambda}\, \xi_n) - \hat{g}_0(\sqrt{\lambda}\, \xi_n)|}{\lambda\, |\xi_n|^2} \right\}$$

If $|\xi_n| \to \infty$, then $|\hat{g}_1(\xi_n)| \leqslant 1/2$ for large n, and $J(\sqrt{\lambda}\, X_0 + \sqrt{1-\lambda}\, X_1) \leqslant (1-\lambda)\, J(X_1) + \lambda/2 J(X_1)$, which is impossible since $J(X_1) \neq 0$. Therefore, extracting a subsequence if necessary, we may assume that $\xi_n \to \xi \in \mathbb{R}^d$. The same argument shows then that $\xi = 0$. But by definition,

$$\frac{|\hat{f}_0(\sqrt{\lambda}\,\xi_n) - \hat{g}_0(\sqrt{\lambda}\,\xi_n)|}{\lambda\,|\xi_n|^2} \to 0$$

as $n \to \infty$. hence, $J(X_\lambda) = 0$, and $J(X_0) = J(X_1) = 0$. ▮

As remarked by Murata and Tanaka[9] and others, a functional having these properties can be used to several applications, as the following.

First Application: The Central Limit Theorem. Let (X_n) be a sequence of independent identically distributed variables with finite variance, and let

$$\xi_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n) \tag{23}$$

Then, if F denotes the common probability law of each X_n, G_n the probability law of ξ_n, and G the gaussian probability law with the same mean vector and covariance matrix as those of F,

$$d_2(G_n, G) \to 0 \qquad \text{as} \quad n \to \infty \tag{24}$$

Moreover, given $\varepsilon > 0$, knowing $d_2(F, G)$ and a modulus of continuity of $D^2\hat{f}$ at the origin, one can compute explicitly n_0 such that for $n \geqslant n_0$, $d_2(G_n, G) \leqslant \varepsilon$.

Proof. For P and Q two probability distributions, let us denote by $P \circ Q$ the rescaled convolution $2^{-d}P(\cdot/\sqrt{2}) * Q(\cdot/\sqrt{2})$. Let $\eta_k = \xi_{2^k}$, and P_k its probability law, then, thanks to Theorem 4,

$$J(\eta_{k+1}) = J(P_k \circ P_k) \leqslant J(P_k) = J(\eta_k):$$

$(J(\eta_k))$ is decreasing, hence converging to some limit l. Admit for a while that there exists Q so that for some subsequence k_p, $d_2(P_{k_p}, Q) \to 0$, so that $l = J(Q)$. Then, since the rescaled convolution is continuous with respect to d_2, $J(P_{k_p} \circ P_{k_p}) \to J(Q \circ Q)$; but it is also $J(P_{k_p+1}) \to l = J(Q)$, so that

$$J(Q \circ Q) = J(Q)$$

This implies that Q is the gaussian distribution G, whence $J(\eta_k) \to 0$. Now, any integer $n \geqslant 1$ can be written $\sum_0^n \alpha_k 2^k$ with $\alpha_k \in \{0, 1\}$, and

$$J(\xi_n) \leqslant \frac{1}{n} \sum \alpha_k 2^k \, J(\eta_k) \to 0$$

Finally, to prove that (P_k) has a weak cluster point with respect to the topology of d_2, it suffices to note that the Fourier transform of F_n is $\hat{f}_n(\xi) = (\hat{f}(\xi/\sqrt{n}))^n$, whose second derivatives can be readily computed,

$$D_{ij}^2 \left(\hat{f}\left(\frac{\xi}{\sqrt{n}}\right)\right)^n = \frac{n-1}{n} \, D_i \hat{f}\left(\frac{\xi}{\sqrt{n}}\right) D_j \hat{f}\left(\frac{\xi}{\sqrt{n}}\right) \hat{f}\left(\frac{\xi}{\sqrt{n}}\right)^{n-2} + \hat{f}\left(\frac{\xi}{\sqrt{n}}\right)^{n-1} D_{ij}^2 \hat{f}\left(\frac{\xi}{\sqrt{n}}\right)$$

If one denotes by ψ a modulus of continuity of $D^2\hat{f}$ near 0, i.e.

$$|D^2\hat{f}(\xi) - D^2\hat{f}(0)| \leqslant \psi(|\xi|)$$

where ψ is chosen to be increasing from 0, we obtain at once that for $D^2\hat{f}_n$ one can take as modulus of continuity

$$\psi_n(t) = \sigma^2 t^2 + \psi\left(\frac{t}{\sqrt{n}}\right)$$

ψ_n is bounded, uniformly in n, by

$$\psi_1(t) = \sigma^2 t^2 + \psi(t)$$

This is enough to conclude.

Stated in this form, this would seem to be only an overcomplicated way of proving the central limit theorem; but the interest of this method is that it can immediately lead to explicit computations. Indeed, let $\varepsilon > 0$, and let us look for n_0 such that for $n \geqslant n_0$, $d_2(G_n, G) \leqslant \varepsilon$. First, since $J(P_k)$ is decreasing, if we look for k_0 such that $J(P_{k_0}) \leqslant \varepsilon$. The proof of Theorem 4 clearly shows that

$$J(P_{k+1}) = J(P_k \circ P_k) \leqslant \sup_{\xi} \left\{ \inf\left(\psi(|\xi|), \left(\frac{1+e^{-|\xi|^2/2}}{2}\right) J(P_k)\right)\right\}$$

Let η such that $\psi(\eta) \leqslant \varepsilon$. As long as $J(P_k \circ P_k) \geqslant \varepsilon$, the supremum can only be obtained for $|\xi| \geqslant \eta$, hence $J(P_{k+1}) \leqslant \mu J(P_k)$ with

$$\mu = \frac{1 + e^{-|\xi|^2/2}}{2} < 1$$

Therefore, one can take

$$k_0 = -\frac{\log J(X_1)}{\log \mu}$$

Now, for $n \geqslant 2^{k_0}$, one writes

$$J(X_n) \leqslant \sum_{k < k_0} \frac{\alpha_k 2^k}{n} J(X_1) + \varepsilon \sum_{k \geqslant k_0} \frac{\alpha_k 2^k}{n} \leqslant \frac{2^{k_0}}{n} J(X_1) + \varepsilon$$

and it suffices to choose $n \geqslant 2^{k_0} J(X_1)/\varepsilon$ for this expression to be less than 2ε. ∎

Second Application: Kac's Theorem. Let X_1 and X_2 be two independent random variables with finite variance, such that

$$\begin{cases} \widetilde{X_1} = X_1 \cos\theta + X_2 \sin\theta \\ \widetilde{X_2} = -X_1 \sin\theta + X_2 \cos\theta \end{cases} \tag{25}$$

are independent for some $\theta \in \mathbb{R} \setminus (\pi/2)\,\mathbb{Z}$. Then X_1 and X_2 are gaussian.

Proof.

$$\begin{aligned} J(\widetilde{X_1}) &\leqslant J(X_1) \cos^2\theta + J(X_2) \sin^2\theta \\ J(\widetilde{X_2}) &\leqslant J(X_1) \sin^2\theta + J(X_2) \cos^2\theta \end{aligned} \tag{26}$$

hence $J(\widetilde{X_1}) + J(\widetilde{X_2}) \leqslant J(X_1) + J(X_2)$. But

$$\begin{cases} X_1 = \widetilde{X_1} \cos\theta - \widetilde{X_2} \sin\theta \\ X_2 = \widetilde{X_1} \sin\theta + \widetilde{X_2} \cos\theta \end{cases} \tag{27}$$

by the same inequality, $J(X_1) + J(X_2) \leqslant J(\widetilde{X_1}) + J(\widetilde{X_2})$, so that there is equality in (26), which implies that X_1 and X_2 are gaussian. ∎

4. APPLICATION TO THE BOLTZMANN EQUATION

In this section, we shall assume $d=3$ for simplicity, but all the results can be generalized readily to any dimension $d \geqslant 2$, or to the onedimensional Kac model. The Boltzmann equation (8) can be studied in weak form for a probability measure as well as for a distribution function,

$$\frac{d}{dt} \int \varphi(v)\, dF(v) = \int_{\mathbb{R}^3 \times \mathbb{R}^3 \times S^2} \sigma\left(\frac{u \cdot n}{|u|}\right) \{\varphi(v') - \varphi(v)\}\, dF(v)\, dF(w)\, dn$$

we shall study this with the conditions

$$\int dF_0(v)\, dv = 1, \qquad \int v\, dF_0(v) = 0, \qquad \int v^2\, dF_0(v) = 3;$$

it is classical that these are preserved under the time-evolution of the Boltzmann equation. Moreover, it is equivalent to use the Fourier transform of the equation:[12]

$$\partial_t \hat{f}(t, \xi) = \int_{S^2} \sigma\left(\frac{\xi \cdot n}{|\xi|}\right) [\hat{f}(\xi^+)\, \hat{f}(\xi^-) - \hat{f}(\xi)\, \hat{f}(0)]\, dn \tag{28}$$

where

$$\begin{cases} \xi^+ = \dfrac{\xi + |\xi|\, n}{2} \\[2ex] \xi^- = \dfrac{\xi - |\xi|\, n}{2} \end{cases} \tag{29}$$

and the initial conditions are such that

$$\hat{f}(0) = 1, \qquad \nabla \hat{f}(0) = 0, \qquad \nabla^2 \hat{f}(0) = -3$$

$\hat{f} \in C^2(\mathbb{R}^d)$. Note that $\xi^+ + \xi^- = \xi$, and $|\xi^+|^2 + |\xi^-|^2 = |\xi|^2$.

Theorem 5. Let F and G be two solutions of the Boltzmann equation (8). Then, for all time $t \geqslant 0$,

$$d_2(F(t), G(t)) \leqslant d_2(F(0), G(0))$$

Before proving Theorem 5, we mention three useful corollaries.

Corollary 5.1. Let F_0 be a nonnegative measure with finite variance. Then, there exists a unique weak solution $F(t)$ of the Boltzmann equation, such that $F(0)=F_0$.

Corollary 5.2. If $F_\varepsilon(t)$ is a sequence of approximate solutions of the Boltzmann equation, obtained by a standard cut-off procedure for instance, then F_ε converges weakly to F. This entails in particular that such results as the decrease of the Fisher information, or the decrease of Tanaka's functional, which are known to hold for the cut-off equation, also hold for the non cut-off equation.

Corollary 5.3. Let F_0 be a nonnegative measure with finite variance, and $F(t)$ the associated solution of the Boltzmann equation. Let M be the Maxwellian distribution with the same mean vector and variance that F_0. Then $d_2(F(t), M)$ is decreasing towards 0.

Proof of Theorem 5. Let F and G be two solutions of the Boltzmann equation, and $\hat{f}$, $\hat{g}$ their Fourier transforms. Then,

$$\partial_t \frac{(\hat{f}-\hat{g})}{|\xi|^2} = \int \sigma\left(\frac{\xi \cdot n}{|\xi|}\right)\left[\frac{\hat{f}(\xi^+)\,\hat{f}(\xi^-)-\hat{g}(\xi^+)\,\hat{g}(\xi^-)}{|\xi|^2} - \frac{\hat{f}(\xi)-\hat{g}(\xi)}{|\xi|^2}\right] dn \tag{30}$$

Now, we do the usual splitting

$$\begin{aligned}
&\left|\frac{\hat{f}(\xi^+)\,\hat{f}(\xi^-)-\hat{g}(\xi^+)\,\hat{g}(\xi^-)}{|\xi|^2}\right| \\
&\quad \leqslant |\hat{f}(\xi^+)| \left|\frac{\hat{f}(\xi)-\hat{g}(\xi)}{|\xi^-|^2}\right| \frac{|\xi^-|^2}{|\xi|^2} + |\hat{g}(\xi^-)| \left|\frac{\hat{f}(\xi^+)-\hat{g}(\xi^+)}{|\xi^+|^2}\right| \frac{|\xi^+|^2}{|\xi|^2} \\
&\quad \leqslant \sup \left|\frac{\hat{f}-\hat{g}}{|\xi|^2}\right| \left(\frac{|\xi^-|^2+|\xi^+|^2}{|\xi|^2}\right) = \sup \left|\frac{\hat{f}-\hat{g}}{|\xi|^2}\right|
\end{aligned}$$

We set

$$h(t, \xi) = \frac{\hat{f}(\xi)-\hat{g}(\xi)}{|\xi|^2}$$

For cut-off molecules, let e be any fixed unit vector and let us denote by

$$S = \int_{S^2} \sigma(n \cdot e)\, dn$$

the total cross-section. By rotational invariance, for all $\xi \neq 0$,

$$\int \sigma\left(\frac{\xi \cdot n}{|\xi|}\right) = S$$

and the preceding computation shows that

$$|\partial_t h - Sh| \leqslant S\, \|h\|_\infty \tag{31}$$

Gronwall's lemma proves at once that for cut-off molecules, $\|h(t)\|_\infty$ is nonincreasing.

Now, let us consider the case of true Maxwell molecules, where $\sigma(v)$ is singular like $(1-v)^{-5/4}$. This singularity corresponds to grazing collisions, i.e., $\xi^+ \sim \xi$, $\xi^- \sim 0$. Since it is nonintegrable, $S = \infty$. Then we split the right-hand side of (28) according to $|1-v| \geqslant \varepsilon$ or $|1-v| < \varepsilon$. For the first term, we use the preceding estimate, while for the other, we use the fact that the singularity is cancelled by the vanishing of $\hat{f}(\xi^+)\, \hat{f}(\xi^-) - \hat{f}(\xi)\, \hat{f}(0)$ for grazing collisions. Indeed, as in ref. 12, let us write

$$\begin{aligned}
&|\hat{f}(\xi^+)\, \hat{f}(\xi^-) - \hat{f}(\xi)\, \hat{f}(0)| \\
&\quad \leqslant |\hat{f}(\xi^+)|\, |\hat{f}(\xi^+) - \hat{f}(\xi)| + |\hat{f}(\xi)|\, |\hat{f}(\xi^-) - \hat{f}(0)| \\
&\quad \leqslant \sup_{|\eta| \leqslant \sup(|\xi|,\, |\xi^+|)} |\nabla \hat{f}(\eta)|\, |\xi^+ - \xi| + \sup_{|\eta| \leqslant |\xi^-|} |\nabla \hat{f}(\eta)|\, |\xi^-|
\end{aligned}$$

Since $|\xi^+|, |\xi^-| \leqslant |\xi|$, $|D^2 \hat{f}(\xi)| \leqslant d$, and $\nabla \hat{f}(0) = 0$, we conclude that

$$|\hat{f}(\xi^+)\, \hat{f}(\xi^-) - \hat{f}(\xi)\, \hat{f}(0)| \leqslant C\, |\xi|\, |\xi^-| \leqslant C\, |\xi|^2\, (1-v)^{1/2}$$

where C depends only on the dimension. This implies that the integrand in (28) is bounded by $C(1-v)^{-3/4}$, and thus the integral is convergent, uniformly in ξ and in t. As a conclusion, setting

$$S_\varepsilon = \int_{S^2} 1_{|1-n\cdot e| \geqslant \varepsilon}\, \sigma(n\cdot e)\, dn$$

$$r_\varepsilon = \sup_{\xi,\, t} \left| \int_{S^2} 1_{|1-(\xi\cdot n)/|\xi|| < \varepsilon}\, \sigma\left(\frac{\xi\cdot n}{|\xi|}\right) [\hat{f}(\xi^+)\, \hat{f}(\xi^-) - \hat{f}(\xi)\, \hat{f}(0)]\, dn \right|$$

we obtain that $r_\varepsilon \to 0$ as $\varepsilon \to 0$, and

$$|\partial_t h(\xi, t) - S_\varepsilon h(\xi, t)| \leqslant S_\varepsilon\, \|h\|_\infty(t) + r_\varepsilon \tag{32}$$

This is equivalent to

$$|\partial_t(h(\xi, t)\, e^{S_\varepsilon t})| \leqslant S_\varepsilon \|h(\cdot, t)\, e^{S_\varepsilon t}\|_\infty + r_\varepsilon e^{S_\varepsilon t}$$

Integrating from 0 to t, we get

$$|h(\xi, t)|\, e^{S_\varepsilon t} \leqslant |h(\xi, 0)| + \int_0^t d\tau (S_\varepsilon \|h(\cdot, \tau)\, e^{S_\varepsilon \tau}\|_\infty + r_\varepsilon e^{S_\varepsilon \tau})$$

Hence, if $H_\varepsilon(t) = \|h(\cdot, t)\, e^{S_\varepsilon t}\|_\infty$,

$$H_\varepsilon(t) \leqslant H_\varepsilon(0) + \int_0^t r_\varepsilon e^{S_\varepsilon \tau}\, d\tau + \int_0^t S_\varepsilon H(\tau)\, d\tau$$

Now, by the generalized Gronwall inequality,

$$u(t) \leqslant \varphi(t) + \int_0^t \lambda(\tau)\, u(\tau)\, d\tau$$

implies

$$u(t) \leqslant \varphi(0) \exp\left\{\int_0^t \lambda(\tau)\, d\tau\right\} + \int_0^t \exp\left\{\int_\tau^t \lambda(\tau)\, d\tau\right\} \frac{d\varphi}{d\tau}\, d\tau$$

Applying this inequality with $\lambda(\tau) = S_\varepsilon$ and $\varphi(t) = H_\varepsilon(0) + \int_0^t r_\varepsilon e^{S_\varepsilon \tau}\, d\tau$, we obtain

$$H_\varepsilon(t) \leqslant H_\varepsilon(0)\, e^{S_\varepsilon t} + t r_\varepsilon e^{S_\varepsilon t}$$

namely

$$\|h(\cdot, t)\|_\infty \leqslant \|h(\cdot, 0)\|_\infty + r_\varepsilon t$$

Letting ε going to 0, we obtain $\|h(\cdot, t)\|_\infty \leqslant \|h(\cdot, 0)\|_\infty$, i.e.

$$d_2(F(t), G(t)) \leqslant d_2(F(0), G(0)) \quad \blacksquare$$

REFERENCES

1. N. M. Blachman. The convolution inequality for entropy powers, *IEEE Trans. Inform. Theory* **2**:267–271 (1965).
2. A. V. Bobylev, G. Toscani. On the generalization of the Boltzmann H-theorem for a spatially homogeneous Maxwell gas, *J. Math. Phys.* **33**:2578–2586 (1992).
3. E. A. Carlen, E. Gabetta, and G. Toscani. Propagation of smoothness and the rate of exponential convergence to equilibrium for a spatially homogeneous Maxwellian gas, *Commun. Math. Phys.* (to appear) (1997).

4. E. A. Carlen, M. C. Carvalho and E. Gabetta. Central limit theorem for Maxwellian molecules and truncation of the Wild expansion, Preprint (1997).
5. I. Csiszar. Information-type measures of difference of probability distributions and indirect observations, *Stud. Sci. Math. Hung.* **2**:299–318 (1967).
6. E. Gabetta, G. Toscani, and B. Wennberg. Metrics for probability distributions and the trend to equilibrium for solutions of the Boltzmann equation, *J. Stat. Phys.* **81**:901–934 (1995).
7. L. Kantorovich. On translation of mass (in Russian) *Dokl. AN SSSR* **37**:227–229 (1942).
8. H. P. McKean, Jr. Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas, *Arch. Rat. Mech. Anal.* **21**:343–367 (1966).
9. H. Murata and H. Tanaka. An inequality for certain functional of multidimensional probability distributions, *Hiroshima Math. J.* **4**:75–81 (1974).
10. R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker–Planck equation, To appear *in SIAM J. Appl. Math. Anal.*
11. Yu. V. Prokhorov. Convergence of random processes and limit theorems in probability theory, *Theory Prob. Appl.* **1**:157–214 (1956).
12. A. Pulvirenti and G. Toscani. The theory of the nonlinear Boltzmann equation for Maxwell molecules in Fourier representation, *Ann. Mat. Pura Appl. 4* **171**:181–204 (1996).
13. A. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon, *Inform. Control* **2**:101–112 (1959).
14. V. Strassen. The existence of probability measures with given marginals, *Ann. Math. Statist.* **36**:423–439 (1965).
15. H. Tanaka. An inequality for a functional of probability distributions and its application to Kac's one-dimensional model of a Maxwellian gas, *Wahrsch. Verw. Geb.* **27**:47–52 (1973).
16. H. Tanaka. Probabilistic treatment of the Boltzmann equation of Maxwellian molecules, *Wahrsch. Verw. Geb.* **46**:67–105 (1978).
17. I. Vaida. *Theory of Statistical Inference and Information*, Kluwer Academic Publishers, Dordrecht (1989).
18. L. N. Vasershtein. Markov processes on countable product space describing large systems of automata (in Russian), *Probl. Pered. Inform.* **5**:64–73, 1969.
19. V. M. Zolotarev. Probability metrics, *Theory Prob. Appl.* **28**:278–302 (1983).