

Convergence of the Iterated Boltzmann Map

By

R.F. STREATER*

Abstract

Consider the Boltzmann equation of degree $d \geq 2$ for a discrete sample space \mathcal{Q} and discrete time; in one step it defines a non-linear map τ on the space of measures on \mathcal{Q} . We show that the entropy increases under τ .

When the energy levels of \mathcal{Q} are equally spaced and scattering conserves energy, the iterated map $\{\tau^m \mu\}_{m=1,2,\dots}$ converges in l_1 to a canonical state μ_β for any initial measure μ , if the scattering matrix mixes each energy shell. If \mathcal{Q} is finite, β depends only on the energy of μ . Under other mixing conditions, $\tau^m \mu$ converges to the microcanonical or grand canonical ensemble.

§ 1. Introduction

1.1 A discrete system can be described by a sample space $\mathcal{Q} = \{0, 1, 2, \dots\}$ together with a probability measure $\mu = (\mu_0, \mu_1, \dots)$ on \mathcal{Q} defining a state of the system. The numbers $\{\mu_i\}$ obey the constraints

$$(1.2) \quad \mu_j \geq 0, \quad j = 0, 1, 2, \dots$$

$$(1.3) \quad \sum_{j=1}^{\infty} \mu_j = 1.$$

We interpret μ_j as the probability that the system in hand is in state j . In this paper we study the time-evolution of μ determined by a non-linear map τ similar to one obtained from the Boltzmann equation. We interpret $\tau \mu$ as the state of the system after one unit of time has passed, if μ is the state at the start of the time-interval. We call τ a Boltzmann map; our main interest is to prove that $\tau^m \mu$ converges in l_1 to an equilibrium state as $m \rightarrow \infty$, and we achieve this under natural conditions on μ and τ . The technique is the usual one:

Communicated by H. Araki, April 15, 1983. Revised November 29, 1983.

* Bedford College, Regents Park, London NW1 4NS, UK. Now at Kings College, London. This work was completed while the author was visiting the Research Institute for Mathematical Sciences, Kyoto University, supported by the Fund for the 70th Anniversary of Kyoto University. I thank Prof. H. Araki for his kind hospitality.

we show that the entropy increases at each step, unless we are at equilibrium; this can then be used to improve on the abstract fact that $\{\tau^m \mu\}_{m=1,2,\dots}$ has a weakly convergent subnet, to show l_1 convergence to equilibrium of the sequence itself. Our techniques can be applied easily to Boltzmann maps of any degree, but we illustrate the method first in the quadratic case, defined as follows.

1.4 Definition. *A quadratic Boltzmann map is a map $\mu \mapsto \tau \mu = \mu'$ of the form*

$$(1.5) \quad \mu'_j = \mu_j + \sum_{k=0}^{\infty} \sum_{\substack{l,m \\ l,m \neq j,k}} T_{jk,lm} (\mu_l \mu_m - \mu_j \mu_k).$$

Here, the numbers $T_{jk;lm}$, called the scattering probabilities obey

$$(1.6) \quad T_{jk,lm} = T_{lm,jk} = T_{kj,ml} \geq 0$$

$$(1.7) \quad \sum_{l,m \neq j,k} T_{jk,lm} \leq 1.$$

Equations of the form (1.5) are used in population dynamics [1,2], especially for species with well-defined generations. Our main interest is in kinetic theory of identical particles, in which case (1.6) expresses symmetry and time-reversal invariance not so natural in population dynamics.

We interpret $\mu_j \mu_k$ as the probability that two samples drawn independently from \mathcal{Q} are respectively in states j and k , so $T_{jk,lm} \mu_j \mu_k$ is the probability that the pair scatter out of that state as $(l, m) \neq (j, k)$. Let us define

$$(1.8) \quad T_{jk,jk} = 1 - \sum_{l,m \neq j,k} T_{jk,lm}.$$

This is the probability of no scattering, and motivates condition (1.7).

Define the entropy of the state μ to be

$$(1.9) \quad S(\mu) = - \sum_j \mu_j \log_e \mu_j$$

when this is finite, with the convention that $0 \log 0 = 0$. If this diverges, we put $S(\mu) = \infty$. Let \mathcal{M} denote the set of probability measures on \mathcal{Q} . The main part of the paper is devoted to proving the theorems below:

Theorem 1. *A quadratic Boltzmann map τ maps \mathcal{M} into itself.*

Theorem 2. *$S(\tau \mu) \geq S(\mu)$ for any $\mu \in \mathcal{M}$.*

Theorem 3. (a) *If \mathcal{Q} is finite and 1 is a simple eigenvalue of the doubly stochastic matrix $T_{jk,lm}$ acting on $\mathcal{Q} \times \mathcal{Q}$, and $\mu \in \mathcal{M}$, then*

$$(1.10) \quad (\tau^m \mu)_j \rightarrow 1/n \quad \text{as } m \rightarrow \infty$$

where $n = |\Omega|$.

(1.11) (b) If each $T_{jk;l m} = 0$ unless $j+k=l+m$, then

$$(1.12) \quad \sum_{j=0}^{\infty} j \mu_j = \sum_{j=0}^{\infty} j (\tau \mu)_j \quad (\text{energy conservation})$$

and if 1 is a simple eigenvalue of the doubly stochastic matrix $(T_{j,N-j;l,N-l})_{j,l=0,1,\dots,N}$ for each N , acting on \mathbb{R}^{N+1} then for any μ with $\sum j \mu_j = \mathcal{E} < \infty$, we have

$$(1.13) \quad \tau^m \mu \rightarrow \hat{\mu} = \left(\frac{e^{-\hat{\beta}j}}{z} \right)_{j=0,1,\dots} \quad \text{in } l_1 \text{ as } m \rightarrow \infty$$

for some $\hat{\beta}$, where $z = \sum_{j \in \Omega} e^{-\hat{\beta}j}$, and $\hat{\beta}$ is a parameter obeying

$$(1.14) \quad \sum j \hat{\mu}_j = \hat{\mathcal{E}} \leq \mathcal{E} = \sum j \mu_j,$$

and when $\mu_0 = 1, \mu_1 = \dots = \mu_j = \dots = 0$ the case $\hat{\mathcal{E}} = 0, \hat{\beta} = \infty$ is understood to mean $\hat{\mu}_0 = 1, \hat{\mu}_1 = \dots = \hat{\mu}_j = \dots = 0$.

If Ω is finite, we have $\hat{\mathcal{E}} = \mathcal{E}$ in (1.14).

1.15 Remark. We interpret j as the energy of state j . We recognise $\hat{\mu}$ as the Gibbs state of temperature $1/\hat{\beta}$ and internal energy $\hat{\mathcal{E}}$. The quantities $\mu_j \mu_k - \mu_l \mu_m$ are called the disequilibrium parameters. Obviously, if they are all zero, then μ is a fixed point of τ . So, $\mu_j = 1/n$ for all j defines a fixed point, the uniform distribution. Theorem 3(a) states that this is the end result of repeated mixings with τ if the doubly stochastic matrix T on $\Omega \times \Omega$ has a simple largest eigenvalue. Indeed, $\mu_j = 1/n$ is the only solution of $\mu_j \mu_k = \mu_l \mu_m$, for all j, k, l, m , and $\sum \mu_j = 1$. In Theorem 3(b) we shall only be able to deduce that $\mu_j \mu_k - \mu_l \mu_m = 0$ on the “two-particle energy shell”, i.e. if $j+k=l+m$, because we assume that T does not mix two-particle states of different total energy. These equations then have a one-parameter family of solutions, $\hat{\mu}(\beta)$. If $|\Omega| = \infty$ we have not been able to prove that the energy of the limit state is \mathcal{E} , even though the energy is \mathcal{E} at every stage of the iteration. Mathematically, this is because $\sum j \mu_j$ is not an l_1 -continuous function of $\mu \in \mathcal{M}$. It is possible that for special choices of initial μ , some energy “escapes up the energy ladder” as we iterate the map. This would correspond to a vanishingly small number of particles carrying away a macroscopic amount of energy, leaving the limit state in equilibrium at a lower temperature. This phenomenon cannot occur if $|\Omega| < \infty$, since then \mathcal{E} is continuous.

We shall prove these theorems in § 2 and § 3, and discuss easy generalizations in § 4. A more general form of the finite case is proved in [2], which is not so easy to generalize to the n-body case as ours.

§ 2. The Finite Case

2.0 *Remark.* The finite case is, of course, a special case of the infinite case proved later; but it is worth while showing how easily the finite case follows from Lemma 3.4, which is assumed in this section. Note that if T satisfies (1.6), then $TT^*=T^2$; this means that if 1 is a simple eigenvalue of T then it is also a simple eigenvalue of TT^* ; thus Lemma 3.4 does apply to such a T .

2.1 *Remark.* We may write the equation as

$$\mu'_j = \mu_j - \sum_k \mu_j \mu_k \sum_{lm \neq j,k} T_{jk;lm} + \sum_k \sum_{lm \neq j,k} \mu_l \mu_m T_{jk;lm};$$

using (1.3) gives

$$\begin{aligned} \mu'_j &= \sum_k \{ \mu_k \mu_j (1 - \sum_{lm \neq j,k} T_{jk,lm}) + \sum_{l,m \neq j,k} \mu_l \mu_m T_{jk;lm} \} \\ &= \sum_k \sum_{l,m} \mu_l \mu_m T_{jk;lm} \end{aligned}$$

using (1.8).

2.2 *Proof of Theorem 1.* We see τ as the composition of three maps. First, $\sigma: \mu \mapsto \mu \otimes \mu$ defines for each μ a probability on $\mathcal{Q} \times \mathcal{Q}$. Then comes the map $\tau_1: \mu_{jk} = \mu_j \mu_k \mapsto \sum_{l,m} T_{jk,lm} \mu_l \mu_m = \mu'_{jk}$ say. Since

$$(2.3) \quad \sum_{jk} T_{jk,lm} = 1$$

τ_1 is a stochastic (indeed, a doubly stochastic) map, and so $\tau_1 \sigma \mu$ is a probability on \mathcal{Q}^2 . Finally, $\tau_2: \mu'_{jk} \mapsto \mu'_j = \sum_k \mu'_{jk}$ takes the marginal probability of μ'_{jk} , and so μ'_j is a probability. This proves Theorem 1.

2.4 *Proof of Theorem 2.* The entropy of $\mu \otimes \mu$ is double the entropy of μ , i.e. $S(\mu \otimes \mu) = 2S(\mu)$. Moreover, a doubly stochastic map does not decrease entropy (Renyi [3], p. 556, but not accurately stated there). Hence $2S(\mu) = S(\sigma \mu) \leq S(\tau_1 \sigma \mu) = S(\mu'_{jk})$. Let $\mu'_j = \sum_k \mu'_{jk}$, $\nu'_k = \sum_j \mu'_{jk}$. Then ([3], p. 557)

$$S(\mu'_{jk}) \leq S(\mu'_j) + S(\nu'_k).$$

But by symmetry, $\mu'_j = \nu'_j$. Thus

$$2S(\mu') = S(\mu'_j) + S(\nu'_j) \geq S(\mu'_{jk}) \geq S(\mu_{jk}) = S(\mu \otimes \mu) = 2S(\mu).$$

This proves theorem 2.

2.5 *Proof of Theorem 3.* Let $|\mathcal{Q}|=n<\infty$. The sequence $(\tau^m\mu)_m\geq 0$ lies in a compact set in \mathbb{R}^n , and so has a convergent subsequence $(\tau^{\lambda_m}\mu)$, converging to $\hat{\mu}$ say. Since $S(\mu)$ is a continuous function,

$$(2.6) \quad \lim_{m\rightarrow\infty} S(\tau^{\lambda_m}\mu) = S(\hat{\mu}).$$

Now τ is a quadratic polynomial, and so is continuous. Hence $\tau\circ\tau^{\lambda_m}\mu\rightarrow\tau\hat{\mu}$ and $S(\tau\circ\tau^{\lambda_m}\mu)\rightarrow S(\tau\hat{\mu})$. But $\{S(\tau^m\mu)\}_{m=0,1,\dots}$ is a bounded monotone sequence, and so converges. Its subsequences $\{S(\tau^{\lambda_m}\mu)\}$ and $\{S(\tau\circ\tau^{\lambda_m}\mu)\}$ therefore converge to the same limit.

Therefore $S(\hat{\mu})=S(\tau\hat{\mu})$.

2.7 *Proof of Theorem 3a.* If 1 is a simple eigenvalue of the stochastic matrix τ_1 , then there is only one fixed point of τ_1 namely $\bar{\mu}$ with $\bar{\mu}_i=j=n^{-2}$. We now show that $\hat{\mu}_i=n^{-1}$. If not, then $\hat{\mu}\otimes\hat{\mu}\neq\bar{\mu}$. So, by Lemma 3.4 below,

$$S(\tau_1(\hat{\mu}\otimes\hat{\mu}))>S(\hat{\mu}\otimes\hat{\mu}).$$

This gives $2S(\tau\hat{\mu})\geq S(\tau_1(\hat{\mu}\otimes\hat{\mu}))>2S(\hat{\mu})=2S(\tau\hat{\mu})$ a contradiction. Thus any convergent subsequence converges to $\hat{\mu}_i=n^{-1}$, so $\tau^m\mu$ must converge to $\hat{\mu}$. This proves (a).

2.8 *Proof of Theorem 3b.* If $T_{j,k,l,m}=0$ unless $j+k=l+m=N$ say, the two-particle energy, then let $\mathcal{E}(\mu)=\sum j\mu_j$ and

$$(2.9) \quad T_{j,l}^N = T_{j,N-j;l,N-l}, \quad 0\leq j, l\leq N.$$

Then

$$\mu'_j = \mu_j + \sum_{N=0}^{\infty} \sum_{l=0}^N T_{j,l}^N (\mu_l\mu_{N-l} - \mu_j\mu_{N-j})$$

and

$$(2.10) \quad \begin{aligned} \sum_0^{\infty} j\mu'_j &= \sum_{j=0}^{\infty} j\mu_j + \sum_{N=0}^{\infty} \delta^N, \quad \text{where} \\ \delta^N &= \sum_{j=0}^N j \sum_{l=0}^N T_{j,l}^N (\mu_l\mu_{N-l} - \mu_j\mu_{N-j}). \end{aligned}$$

Replace j by $N-j$ and l by $N-l$ in the sums. Then

$$\delta^N = \sum_{j=0}^N (N-j) \sum_{l=0}^N T_{N-j,N-l}^N (\mu_{N-l}\mu_l - \mu_{N-j}\mu_j).$$

Use the symmetry $T_{j,l}^N = T_{N-j,N-l}^N$, so

$$(2.11) \quad \delta^N = \sum_{j=0}^N (N-j) \sum_{l=0}^N T_{j,l}^N (\mu_l \mu_{N-l} - \mu_j \mu_{N-j}).$$

Add (2.10) and (2.11), giving

$$2\delta^N = \sum_{j=0}^N N \sum_{l=0}^N T_{j,l}^N (\mu_l \mu_{N-l} - \mu_j \mu_{N-j}).$$

Use time reversal invariance $T_{j,l}^N = T_{l,j}^N$ giving $\delta^N = 0$.

Hence $\mathcal{E}(\tau\mu) = \mathcal{E}(\mu)$: energy is conserved.

Since \mathcal{E} is a continuous function of μ , $\mathcal{E}(\hat{\mu}) = \lim \mathcal{E}(\tau^m \mu) = \lim \mathcal{E} = \mathcal{E}$.

We now show as in (a) that $\hat{\mu} \otimes \hat{\mu}$ is the equidistribution on each energy shell. If not, Lemma 3.4 gives

$$2S(\tau\hat{\mu}) \geq S(\tau_1(\hat{\mu} \otimes \hat{\mu})) > S(\hat{\mu} \otimes \hat{\mu}) = 2S(\hat{\mu}) = 2S(\tau\hat{\mu}),$$

a contradiction. Hence all “energy-conserving” disequilibrium parameters are equal: $\hat{\mu}_j \hat{\mu}_k = \hat{\mu}_l \hat{\mu}_m$ if $j+k=l+m$. The only solution to this, with $\mathcal{E}(\hat{\mu}) = \mathcal{E}$, is

$$\hat{\mu}_j = \frac{e^{-\beta j}}{Z}, \quad \beta \in \mathbf{R} \cup \infty \quad \text{with}$$

$$Z = \sum_{j \in \mathcal{Q}} e^{-\beta j}$$

and β determined by $\sum_{j \in \mathcal{Q}} j \hat{\mu}_j = \mathcal{E} = \sum j \mu_j$.

This is the only state with this entropy (and having energy \mathcal{E}), so all convergent subsequences converge to $\hat{\mu}$, and as in (a), $\tau^m \mu \rightarrow \hat{\mu}$ as $m \rightarrow \infty$. This proves 1.14 and Theorem 3 when $|\mathcal{Q}| < \infty$.

§ 3. The Infinite Case

We shall use the following results.

3.1. Let $g \in C^2(\mathbf{R})$, $x = \sum_{k=1}^n w_k x_k$, $\sum w_k = 1$, $w_k \geq 0$.

Then by Taylor’s theorem

$$(3.2) \quad g(x_k) = g(x) + (x_k - x)g'(x) + \frac{1}{2}(x_k - x)^2 g''(\xi_k)$$

where $x \leq \xi_k \leq x_k$ or $x_k \leq \xi_k \leq x$.

Therefore summing $w_k g(x_k)$ gives

$$\sum w_k g(x_k) = g(x) + \frac{1}{2} \sum_k w_k (x - x_k)^2 g''(\xi_k).$$

Apply this to $g(x)=x \log x$ where $0 \leq x_k \leq 1$:

$$(3.3) \quad \sum w_k x_k \log x_k = x \log x + \frac{1}{2} \sum_k w_k (x - x_k)^2 \xi_k^{-1} \\ \geq x \log x + \frac{1}{2} \sum_k w_k (x - x_k)^2 \quad \text{as } \xi_k \leq 1.$$

3.4 Lemma (gain in entropy under a doubly stochastic map). Let T be a doubly stochastic map on \mathbb{R}^N and let p be a semi-probability, i.e. $1 \geq p_j \geq 0$, $j=1, \dots, N$. Let the eigenvalue 1 of TT^* be simple, with a gap Δ to the next largest eigenvalue. Let $q_j = \sum_k T_{kj} p_k$. Then

$$(3.5) \quad \sum_k p_k \log p_k \geq \sum_j q_j \log q_j + \frac{1}{2} \Delta \|p - \bar{p}\|_2^2$$

where $\bar{p}_j = N^{-1} \|p\|_1$.

Proof. Since $T_{kj} \geq 0$ and $\sum_k T_{kj} = 1$, we may apply (3.3) for each j to give, as $p_j \leq 1$,

$$\sum_k T_{kj} p_k \log p_k \geq q_j \log q_j + \frac{1}{2} \sum_k T_{kj} (q_j - p_k)^2.$$

Summing over j gives, as $\sum_j T_{kj} = 1$:

$$\begin{aligned} \sum_k p_k \log p_k - \sum_j q_j \log q_j &\geq \frac{1}{2} \sum_j \sum_k T_{kj} (q_j - p_k)^2 \\ &= \frac{1}{2} (\sum_j q_j^2 - 2 \sum_j \sum_k T_{kj} p_k q_j + \sum_k p_k^2) \\ &= \frac{1}{2} (\|T^* p\|_2^2 - 2 \langle T^* p, T^* p \rangle + \|p\|_2^2) \\ &= \frac{1}{2} \langle (1 - TT^*) p, p \rangle = \frac{1}{2} \langle (1 - TT^*) (p - \bar{p}), (p - \bar{p}) \rangle \end{aligned}$$

since $(1 - TT^*) \bar{p} = 0$

$$\geq \frac{1}{2} \Delta \|p - \bar{p}\|_2^2$$

since $p = \bar{p} \oplus (p - \bar{p})$ and Δ is the smallest eigenvalue of $(1 - TT^*)$ on the space orthogonal to \bar{p} .

This proves the lemma.

3.6 Remark. This shows that if $\Delta > 0$, $T^n p \rightarrow \bar{p}$ as $n \rightarrow \infty$; for entropy increases and is bounded, so $\|p_n - \bar{p}\|_2^2 \rightarrow 0$ where $p_n = T^n p$.

3.7. Now return to the Boltzmann map τ and consider the entropy of $\tau^m\mu$. This state has energy $\mathcal{E} = \sum j\mu_j$, and therefore its entropy is bounded by the entropy of the thermodynamic state of this energy, which is given by

$$(3.8) \quad S(\mathcal{E}) = -\sum e^{-j\beta} Z^{-1}(\beta) \log(Z^{-1}e^{-j\beta}) = \beta\mathcal{E} + \log Z(\beta)$$

where β is determined by

$$(3.9) \quad \mathcal{E} = Z^{-1} \sum j e^{-j\beta},$$

$$(3.10) \quad Z = \sum e^{-j\beta} = (1 - e^{-\beta})^{-1}.$$

Hence $S(\tau^m\mu)$ is an increasing bounded sequence and so converges as $m \rightarrow \infty$. The increments $S(\tau^{m+1}\mu) - S(\tau^m\mu)$ therefore converge to zero, which a fortiori implies that the gain in entropy under τ_1 , applied to each energy shell N of $\tau^m\mu \otimes \tau^m\mu$, goes to zero as $m \rightarrow \infty$. This map τ_1 , restricted to the energy shell $j+k=N$, is a doubly stochastic map with a spectral gap, and so Lemma 3.4 applies with $N+1$ points and the semi-probability $p = (p_1, \dots, p_{N+1})$ where

$$p_j = (\tau^m\mu)_{j-1}(\tau^m\mu)_{N-j+1}.$$

Since $\Delta > 0$ for each N , $p_j - \bar{p} \rightarrow 0$ as $m \rightarrow \infty$ which implies that the disequilibrium parameters

$$p_j - p_k = (\tau^m\mu)_{j-1}(\tau^m\mu)_{N-j+1} - (\tau^m\mu)_{k-1}(\tau^m\mu)_{N-k+1} \rightarrow 0$$

for each N , as $m \rightarrow \infty$. That is, all the energy-conserving disequilibrium parameters of $\tau^m\mu$ converge to zero.

3.11. Since the set of probabilities lies in a weakly compact set in $l_1(\mathcal{Q})$, a subnet say $\{\nu\}$, of $\{\tau^m\mu\}$ converges weakly to $\hat{\mu}$ say, i.e. $\nu_j \rightarrow \hat{\mu}_j$. Since the disequilibrium parameters of $\{\tau^m\mu\}$ converge to zero, they converge to zero also along the subnet. Thus $\hat{\mu}_j\hat{\mu}_k - \hat{\mu}_l\hat{\mu}_m = \lim(\nu_j\nu_k - \nu_l\nu_m) = 0$ if $j+k=l+m$. Hence there are parameters $\hat{Z}, \hat{\beta}$ such that

$$(3.12) \quad \hat{\mu}_j = \hat{Z}^{-1} e^{-j\hat{\beta}}.$$

Obviously for any N ,

$$(3.13) \quad \sum_{j=0}^N \nu_j \leq 1, \quad \sum_{j=0}^N j\nu_j \leq \mathcal{E},$$

and so the limits (along the subnet) obey

$$\sum_{j=0}^N \hat{\mu}_j \leq 1, \quad \sum_{j=0}^N j\hat{\mu}_j \leq \mathcal{E}, \quad \text{for any } N.$$

Hence

$$(3.14) \quad \sum_{j=0}^{\infty} \hat{\mu}_j \leq 1, \quad \hat{\mathcal{E}} = \sum_{j=0}^{\infty} j \hat{\mu}_j \leq \mathcal{E}.$$

3.15 We now show that $\sum \hat{\mu}_j = 1$, i.e. no probability is squeezed out in the limit. Choose $\epsilon > 0$ and N large enough so that $\mathcal{E}/N < \epsilon/2$; then proceed along the subnet up to ν such that

$$|\nu_0 + \dots + \nu_N - \hat{\mu}_0 - \dots - \hat{\mu}_N| < \frac{\epsilon}{2}.$$

Now for all N and ν , $\sum_{j=0}^{\infty} j \nu_j \leq \mathcal{E}$ and so $\sum_{j=0}^{\infty} \nu_j \leq \frac{\mathcal{E}}{N+1} < \frac{\epsilon}{2}$. Hence

$$\begin{aligned} |1 - \sum_{j=0}^N \hat{\mu}_j| &= |\nu_0 + \dots - \hat{\mu}_0 - \dots - \hat{\mu}_N| \leq \left| \sum_{j=0}^N \nu_j - \sum_{j=0}^N \hat{\mu}_j \right| + \sum_{j=N+1}^{\infty} \nu_j \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2}. \end{aligned}$$

Hence $\sum_0^N \hat{\mu}_j \rightarrow 1$ as $N \rightarrow \infty$. This, together with (3.12), says that $\hat{\mu}$ is a Gibbs state of temperature $1/\hat{\beta}$.

3.16. We now show that $\beta \leq \hat{\beta}$. Indeed

$$\frac{\partial \mathcal{E}(\beta)}{\partial \beta} = \frac{\partial}{\partial \beta} (\sum j e^{-j\beta} Z^{-1}(\beta)) = -\mathbb{E}J^2 + (\mathbb{E}J)^2 \leq 0$$

where J is the random variable $J(j) = j$ and \mathbb{E} is expectation in the Gibbs state of temperature β^{-1} . Since $\mathcal{E}(\hat{\beta}) = \hat{\mathcal{E}} \leq \mathcal{E} = \mathcal{E}(\beta)$, it follows that $\hat{\beta} \geq \beta$.

3.17. We now show that $S(\hat{\beta}) \leq S(\beta)$, where $S(\beta)$ is given by (3.8). We see

$$\frac{\partial S}{\partial \beta} = \mathcal{E} + \beta \frac{\partial \mathcal{E}}{\partial \beta} + \frac{\partial}{\partial \beta} (\log Z(\beta)) = \beta \frac{\partial \mathcal{E}}{\partial \beta} \leq 0.$$

Since $\hat{\beta} \geq \beta$ it follows that $S(\hat{\beta}) \leq S(\beta)$.

3.18. We now show that $S(\nu) \rightarrow S(\hat{\beta})$ along the subnet. First we give an elementary result.

3.19 Lemma. Let p_n, p_{n+1}, \dots be a sequence of non-negative numbers, with $\sum_n p_j = q$ and $\sum_n j p_j = E$. Let $s = -\sum_n p_j \log p_j$. Then

$$(3.20) \quad s \leq (y+q) \log (y+q) - y \log y - 2q \log q$$

where $y = E - nq$.

For the proof, we maximise the function $s(p)$ subject to the constraints by the method of Lagrange multipliers. The right-hand side of (3.20) is the result.

3.21. By the mean value theorem, $(y+q) \log (y+q) = y \log y + q \frac{\partial}{\partial \xi} (\xi \log \xi)$,

$$y \leq \xi \leq y+q \leq E.$$

Hence $s \leq -2q \log q + q(1 + \log \xi) \leq -2q \log q + q(1 + \log E)$. This result means that a small tail of a distribution, with q small, cannot contain much entropy. This fact allows us to show that $S(\nu)$ increases to $S(\lambda)$ along the subnet.

3.22 **Theorem.** $S(\nu) \rightarrow S(\lambda)$ along the subnet.

Proof. Choose $\epsilon > 0$. Let $q > 0$ be such that

$$-2q \log q + q(1 + \log \mathcal{E}) < \frac{\epsilon}{3}.$$

Let N be chosen large enough so that

$$\sum_{N+1}^{\infty} -\lambda_j \log \lambda_j < \frac{\epsilon}{3}, \quad \sum_{N+1}^{\infty} \lambda_j < \frac{q}{2}.$$

With this N , follow the subnet until each ν_1, \dots, ν_N is so close to the limit that

$$\left| \sum_0^N \nu_j \log \nu_j - \sum_0^N \lambda_j \log \lambda_j \right| < \frac{\epsilon}{3} \quad \text{and}$$

$$\left| \sum_0^N \nu_j - \sum_0^N \lambda_j \right| < \frac{q}{2}$$

hold for all ν further down the subnet. Then

$$\begin{aligned} \left| 1 - \sum \nu_j \right| &= \left| \left(1 - \sum \lambda_j \right) - \left(1 - \sum \nu_j \right) - \sum_{N+1}^{\infty} \lambda_j \right| \\ &< \left| \sum \lambda_j - \sum \nu_j \right| + \left| \sum_{N+1}^{\infty} \lambda_j \right| \\ &< \frac{q}{2} + \frac{q}{2} = q, \quad \text{for all further } \nu. \end{aligned}$$

Hence

$$\begin{aligned} -\sum_{N+1}^{\infty} \nu_j \log \nu_j &\leq -2q \log q + q(1 + \log \mathcal{E}) \\ &\leq \frac{\epsilon}{3}. \end{aligned}$$

Therefore

$$\begin{aligned}
 |-\sum_0^\infty \nu_j \log \nu_j + \sum_0^\infty \hat{\mu}_j \log \hat{\mu}_j| &= |\sum_0^N \hat{\mu}_j \log \hat{\mu}_j - \sum_0^N \nu_j \log \nu_j \\
 &\quad + \sum_{N+1}^\infty \hat{\mu}_j \log \hat{\mu}_j - \sum_{N+1}^\infty \nu_j \log \nu_j| \\
 &\leq |\sum_0^N (\hat{\mu}_j \log \hat{\mu}_j - \nu_j \log \nu_j)| + |\sum_{N+1}^\infty \hat{\mu}_j \log \hat{\mu}_j| + |\sum_{N+1}^\infty \nu_j \log \nu_j| \\
 &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3}.
 \end{aligned}$$

Since this holds for all ν further on, we have $S(\nu) \rightarrow S(\hat{\mu})$ along the subnet.

3.23 *Uniqueness of the limit.*

If $\{\nu\}$ is any convergent subnet of $\{\tau^m \mu\}$, the arguments given in (3.11) show that ν converges elementwise to some Gibbs state $\mu(\beta_1)$ and subsequent sections show that $S(\nu) \rightarrow S(\beta_1)$. But $S(\tau^m \mu)$ converges, so it and all convergent subnets converge to the same limit. Hence $S(\beta_1) = S(\hat{\beta})$. But $\mu(\beta_1)$ and $\mu(\hat{\beta})$ are Gibbs states, and for these, equality of entropy implies equality. Hence $\beta_1 = \hat{\beta}$ and all convergent subnets have the same limit.

3.24 *Proof of Theorem 3b, infinite case.* Suppose $(\tau^m \mu)_j$ does not converge to $\hat{\mu}_j$ for some j . Then there is $\epsilon > 0$ such that the subset of m with $|(\tau^m \mu)_j - \hat{\mu}_j| > \epsilon$ is infinite, and so $\tau^m \mu$ will have a convergent subnet not converging to $\hat{\mu}_j$, contradicting 3.23. Therefore $(\tau^m \mu)_j \rightarrow \hat{\mu}_j$. To show l_1 -convergence, choose $\epsilon > 0$ and N such that $\sum_N^\infty \hat{\mu}_j < \epsilon/2$. Choose M large enough so that for $m > M$

$$|(\tau^m \mu)_0 - \hat{\mu}_0| < \frac{\epsilon}{4}, \quad |(\tau^m \mu)_1 - \hat{\mu}_1| < \frac{\epsilon}{8}, \quad \dots \quad |(\tau^m \mu)_{N-1} - \hat{\mu}_{N-1}| < \frac{\epsilon}{2^{N+1}}.$$

Then for $m > M$

$$\begin{aligned}
 |1 - \sum_{j=0}^{N-1} (\tau^m \mu)_j| &= |1 - \sum_{j=0}^{N-1} [(\tau^m \mu)_j - \hat{\mu}_j] - \sum_0^{N-1} \hat{\mu}_j| \\
 &\leq |1 - \sum_0^{N-1} \hat{\mu}_j| + |(\tau^m \mu)_0 - \hat{\mu}_0| + \dots + |(\tau^m \mu)_{N-1} - \hat{\mu}_{N-1}| \\
 &\leq \frac{\epsilon}{2} + \frac{\epsilon}{4} \left(1 + \frac{1}{2} + \frac{1}{4} + \dots\right) < \epsilon.
 \end{aligned}$$

Then

$$\begin{aligned}
 \sum_1^\infty |(\tau^m \mu)_j - \hat{\mu}_j| &\leq \sum_0^{N-1} |(\tau^m \mu)_j - \hat{\mu}_j| + \sum_N^\infty |(\tau^m \mu)_j - \hat{\mu}_j| \\
 &< \frac{\epsilon}{2} + \sum_N^\infty (\tau^m \mu)_j + \sum_N^\infty \hat{\mu}_j \leq \frac{\epsilon}{2} + \epsilon + \frac{\epsilon}{2}.
 \end{aligned}$$

This proves the theorem.

3.25. We remark that while $\beta < 0$ is possible in the finite case, β must be > 0 in the infinite case.

§ 4. Generalizations

The method gives results for a Boltzmann equation of any degree d , without much extra work. We just consider the map $\sigma: \mu \rightarrow \mu \otimes \mu \otimes \dots \mu$, taking a measure μ on \mathcal{Q} to the product measure on \mathcal{Q}^d . Then the map τ_1 of § 2 is generalized to a doubly stochastic map on the space of measures on \mathcal{Q}^d , of the form

$$(4.1) \quad \mu'_{j_1 \dots j_d} = (\tau_1 \sigma \mu)_{j_1 \dots j_d} = T_{j_1 \dots j_d, k_1 \dots k_d} (\sigma \mu)_{k_1 \dots k_d}$$

(summation convention) for some parameters T , abbreviated to $T_{j,k}$, where $j=(j_1, \dots, j_d)$, $k=(k_1, \dots, k_d)$. We assume that T obeys time-reversal symmetry

$$(4.2) \quad T_{j,k} = T_{k,j}$$

and the physically natural symmetries

$$(4.3) \quad T_{j,k} = T_{\pi(j), \pi(k)}$$

where $\pi \in S_d$ is the permutation

$$\pi(j) = (j_{\pi(1)}, j_{\pi(2)}, \dots, j_{\pi(d)}).$$

The generalized τ_2 is the marginal map

$$\mu'_j = \sum_{j_2, \dots, j_d} \mu'_{j j_2 \dots j_d} = (\tau_2 \tau_1 \sigma \mu)_j.$$

Then the generalized Theorem 1 is immediate, namely, μ' is a probability.

Eq. (4.2) ensures that T is doubly stochastic. Repeated application of the method of (2.4) gives the generalized Theorem 2, namely $S(\tau \mu) \geq S(\mu)$.

To proceed to Theorem 3 we must make some assumption about the mixing properties of τ_1 . If \mathcal{Q} is finite and τ_1 has a unique fixed point, then we get the generalized Theorem 3(a). Now suppose τ_1 conserves energy, i.e.

$$(4.4) \quad T_{j,h} = 0 \quad \text{unless} \quad \sum_{\omega=1}^d j_\omega = \sum_{\omega=1}^d k_\omega.$$

Then it follows by an argument just like (2.10), (2.11) that

$$(4.5) \quad \sum_0^\infty j \mu_j = \sum_0^\infty j(\tau \mu)_j$$

holds, provided that (4.2) and (4.3) hold. Suppose that τ_1 , restricted to the energy shell, $\sum j_\alpha = \sum k_\alpha = N$, has a unique fixed point. Then, as before, all the disequilibrium constants (now of degree d) converge to zero along a subnet. Among the disequilibrium constants will be

$$\mu_{\alpha_1} \cdots \mu_{\alpha_{d-2}} (\mu_j \mu_k - \mu_l \mu_m) \quad \text{with } j+k = l+m.$$

Summing over $\alpha_1, \dots, \alpha_{d-2}$ just gives $\mu_j \mu_k - \mu_l \mu_m \rightarrow 0$ along the subnet in the finite case, and from here on, the proof is the same as in the quadratic case, $d=2$.

In the infinite case, the argument of 3.15 shows that not all μ_j converge to zero, so along a subnet one of the μ_ω is bounded away from 0. This then implies that $\mu_j \mu_k - \mu_l \mu_m \rightarrow 0$ if $j+k=l+m$, and we follow the argument of §3 again to get Theorem 3(b).

4.6. On physical grounds we would expect $T_{j_1 \dots j_d, k_1, \dots, k_d}$ to exhibit "cluster" properties, in that $T_{j_1, j_2, j_3 \dots j_d, k_1, k_2, j_3 \dots j_d}$ should be the probability of the two-particle scattering, $j_1 + j_2 \rightarrow k_1 + k_2$, the remaining particles acting as "spectators"; the same remarks apply to all the other possible subchannels. Only the term with all j_ω 's differing from the k_ω 's gives an intrinsic d -body term in the Boltzmann equation. The term with spectators can be collected together using $\sum \mu_j = 1$ to lower the degree of the scattering term, which can then be made to correspond to a term in the Boltzmann equation of degree $< d$. The degree d of the Boltzmann equation is thus not homogeneous: only the intrinsic d -body term cannot be reduced in degree. We saw this phenomenon for $d=2$, in that (1.5), containing linear and quadratic terms, is equivalent to the apparently homogeneous equation in 2.1.

With this remark we see that the Boltzmann equation with larger d is a real generalization of the equation with smaller d rather than a different equation.

4.7 If \mathcal{Q} is described by a double sequence (j, k) , $j=0, 1, 2, \dots$; $k=0, 1, 2, \dots$, then we can use j to describe one random variable, energy say, and k to describe another, such as particle number, if that can vary in the given region by scattering into another, for example. We can model the exchange of energy and particles between two regions by considering a scattering matrix T , on measures

$\mu \otimes \mu$ (on \mathcal{Q}^2), which is zero unless $j_1 + j_2 = j_3 + j_4$ and $k_1 + k_2 = k_3 + k_4$ where $T = T_{j_1 k_1, j_2 k_2; j_3 k_3, j_4 k_4}$.

These two conditions express the conservation of energy and particle number, respectively. Following the same arguments as in § 2 or § 3, modified suitably, we find that the map $\tau = \tau_2 \tau_1 \sigma$ increases entropy except at a fixed point, and mixes the probabilities on the energy-number shell $j_1 + j_2 = E, k_1 + k_2 = N$. The vanishing of the disequilibrium parameters obeying these conditions implies that μ_{jk} can be expressed as a grand canonical ensemble

$$\mu_{jk} = Z^{-1} e^{-\beta(j - \gamma k)}, \quad \beta, \gamma \in \mathbb{R}$$

which is therefore the limit of $\tau^m \mu$ as $m \rightarrow \infty$. The parameters β (inverse temperature) and γ (chemical potential) are determined by the internal energy and the density

$$\mathcal{E} = \sum_{j,k=0}^{\infty} j \mu_{jk} \quad \rho = \sum_{j,k=0}^{\infty} k \mu_{jk}$$

both of which are conserved by the map $\lim \tau^m$, at least in the finite case. In the infinite case, we have again been unable to exclude the possibility of energy or number being lost to ∞ .

§ 5. Final Remarks

5.1. We have been able to set up an analogous theory for the quantum Boltzmann equation [4]. Convergence is proved provided that the multiplicity $m(j)$ of the j^{th} energy-level is bounded by a polynomial:

$$m(j) \leq \kappa j^r, \quad j = 1, 2, \dots$$

for some $\kappa > 0$ and integer r . This result includes the commutative (classical) case, and so gives a generalization of Theorem 3 to the case with multiplicity.

5.2. A map τ of the type in this paper can be used to construct an associated stochastic process for each initial probability μ . See [5], [6] for a similar but much deeper problem with continuous time. In our case we have $\mathcal{Q} = \{0, 1, 2, \dots\}$ and we seek a probability measure M on $\mathcal{Q} \times \mathcal{Q} \times \dots$ such that the marginal probabilities of M coincide with the probabilities $(\tau^m \mu)$, $m = 0, 1, 2, \dots$. Explicitly, consider the event

$$(5.3) \quad A^m(i) = \{\omega = (\omega_1, \omega_2, \dots) \in \mathcal{Q} \times \mathcal{Q} \times \dots \mid \omega_m = i\}.$$

Then we require M to satisfy

$$(5.4) \quad M(A^m(i)) = (\tau^m \mu)_i, \quad i \geq 0, m \geq 0.$$

This can be achieved if we introduce the symbol

$$(5.5) \quad M^m(i, k) = \sum_{j,l} T_{ij;kl}(\tau^{m-1} \mu)_l.$$

As in the linear case, i.e. Markov chains, this gives us an algorithm for defining M , namely

$$(5.6) \quad M\{\omega: \omega_0 = i_0, \omega_1 = i_1, \dots, \omega_m = i_m\} = M^1(i_1, i_0)M^2(i_2, i_1) \dots M^m(i_m, i_{m-1}).$$

Then it can be shown that M is a probability measure on the cylinder sets of $\mathcal{Q} \times \mathcal{Q} \times \dots$, with the correct marginal probabilities (5.4) and obeying the Markov condition

$$(5.7) \quad M\{A^m(i) | \omega: \omega_0 = i_0, \dots, \omega_{m-1} = i_{m-1}\} = M(A^m(i) | A^{m-1}(i_{m-1})) \\ = M^m(i, i_{m-1}),$$

showing that $M^m(i, k)$ is in fact the conditional probability $M(A^m(i) | A^{m-1}(k))$.

References

- [1] Moran, P.A.P., Entropy, Markov processes and Boltzmann's H-theorem. *Proc. Camb. Phil. Soc.* 57 (1961), 833-842.
- [2] Nishimura, S., Random Collision Processes and their Limiting Distribution using the Discrimination Information. *J. Appl. Prob.*, (1974), 266-280.
- [3] Renyi, A., *Probability Theory*, North Holland, 1970.
- [4] Streater, R.F., Convergence of the Quantum Boltzmann Map, Submitted to *Commun. Math. Phys.*
- [5] McKean, H.P., A Class of Markov Processes Associated with Nonlinear Parabolic Equations. *Proc. Nat. Acad. Sci. (USA)*, 56 (1966), 1907-1911.
- [6] Tanaka, H., On the Uniqueness of Markov Process Associated with the Boltzmann Equation of Maxwellian Molecules, *Proc. Intern. Symp. SDE, Kyoto 1976* (K. Ito, Ed.), Wiley Interscience, (1978) 409-425.

