LECTURES ON STATISTICAL MECHANICS (II)

Notes by J.E. Marsden,
based on lectures by A.S. Wightman.

Fine Hall,

Princeton University,

Princeton, New Jersey, 08540

PREFACE

These notes follow the lectures except for a small number of modifications, especially in §6,9,10.

Although no new results are given (except for a few technicalities), the notes should prove helpful as a guide to what has been, and what currently is being done in the subject. Consequently, the treatment is more complete for the well established aspects of the subject (ch. I,II), and is sketchy for the remainder (ch. III). The emphasis is on the mathematical foundations of the subject, and a precise treatment of theorems.

A certain amount of mathematical maturity on the part of the reader is helpful, as well as a nodding acquaintance with statistical mechanics. For background and further discussion, we refer the reader to Huang [1], Arnold-Avez [1] and Abraham [1], (Bibliographies are at the end of each chapter).

There are a number of errors in these notes although most of them are mere oversights. For example on page 46, the figure is incorrect. We would appreciate knowing of errors, however minor.

J. E. Marsden

A. S. Wightman

July, 1967

<u>Remark added in proof</u>. Lemma 9.16 is incorrect. The counter example is provided by $\sin(nx)$. (Riemann-Lebesque lemma). The correct conclusion is: if $g_n \longrightarrow g$ in the sense of distributions and $\epsilon > 0$, A is a set of positive measure, $|g_n - g| < \epsilon$ for some $x \in A$.

Thus wherever 9.16 is applied to give convergence a.e., use 9.15 if it applies, or else conclude convergence in the sense of distributions. (e.g. in 9.17).

CONTENTS.

CHAPTER I ERGODIC THEORY OF DYNAMICAL SYSTEMS.

This chapter is a condensation and review of some of last year's lectures, with a few modifications and additions. The basic reference here is Arnold-Avez [1]. Only some of the proofs are given (the easy ones) so that the serious student should supplement the treatment here.

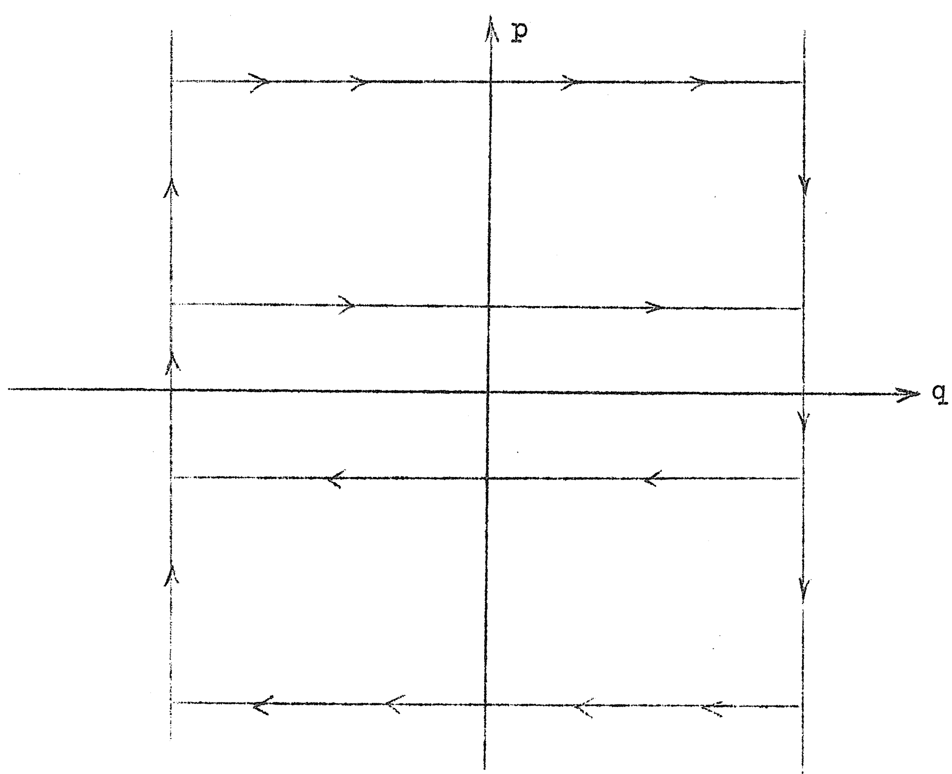
The bulk of the chapter is devoted to a systematic study of flows, or dynamical systems and their ergodic properties of relevance in statistical mechanics. The most recent work of Sinai, Anosov and others is described in §5. In §6 we prove the virial theorem in a modern, slightly generalized form, and briefly discuss the unsolved problems associated with transport properties (viscosity etc.).

§1. Dynamical Systems.

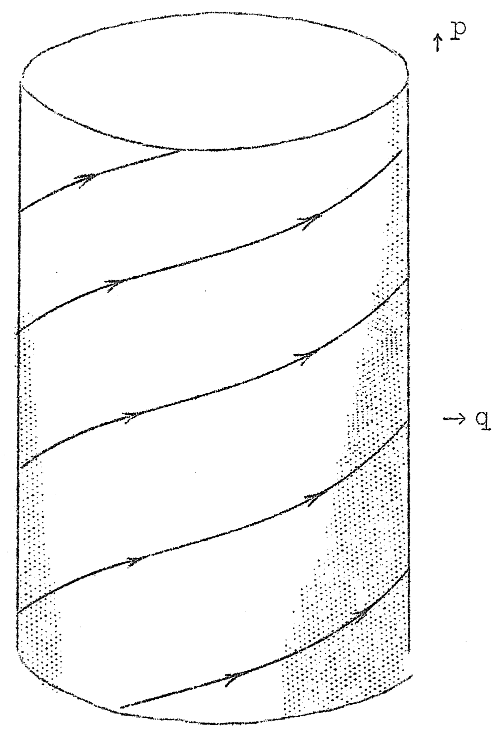
This section summarizes the basic facts about dynamical systems, or flows. Proofs are omitted except where they are trivial, but most definitions are included, so the exposition is essentially self contained. References for this section are Arnold-Avez [1], Halmos [1] and Billingsly [1].

Intuitively, we think of a flow as the time development of the points of a set X, which represent the states of a physical system.

Systems possess a variety of smoothness conditions, so there is a number of conditions we shall impose on the flow. For example, the motion of a particle in a box has a discontinuous flow on its phase space as follows:



On the other hand, a classical mechanical system with a smooth Hamiltonian has a smooth flow. (The configuration space in this example is a circle, the phase space a cylinder.)



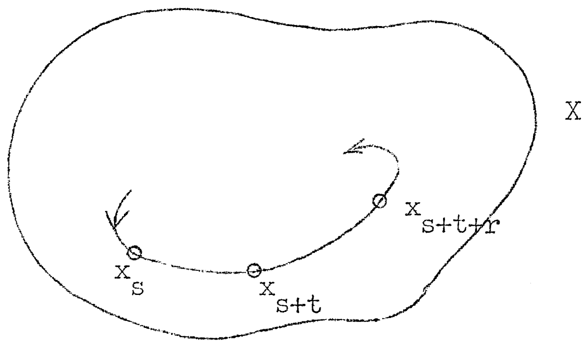
The states at a time t are obtained from those at a time s by a mapping

$$T(t,s): \quad X \longrightarrow X$$

which, if the process is deterministic should obey

$$T(t,s)\circ T(s,r) = T(t,r)$$

and $T(t,t)$ is the identity map $(T(t,t)x=x)$.



These notions are made precise as follows.

1.1 Definition. Let X be a set. A flow (or dynamical system) on X is a mapping

$$T:\ R \times R \times X \longrightarrow X$$

(where R denotes the reals and $Y \times Z = \{(y,z):\ y \in Y \text{ and } z \in Z\}$ denotes the Cartesian product) such that $T(t,s):\ X \longrightarrow X$ defined by $T(t,s)(x) = T(t,s,x)$ satisfies

(i) $T(t,s)\circ T(s,r) = T(t,r)$ and

(ii) $T(t,t)$ is the identity; for all $t,\ s,\ r \in R$, where $\circ$ denotes composition.

Let $R^+ = \{t \in R:\ t \geq 0\}$ and Z denote the integers. A mapping $T:\ R^+ \times R^+ \times X \longrightarrow X$ is called a one sided flow iff T satisfies (i) and (ii) above. Similarly, replacing R by Z defines a cascade, and replacing $R \times R$ by $\{(t,s):\ t \geq s\}$ gives an irreversible flow (or cascade).

A flow is called stationary iff

$$T(t+u,\ s+u) = T(t,s)$$

for all $t, s, u \in R$. Similarly for irreversible and cascades.

The orbit of $x \in X$ under a flow (or cascade) T is the set

$$\{T(t,o)\cdot x:\quad t \in R\}$$

and the positive orbit is the set

$$\{T(t,o)\cdot x:\quad t \in R^+\}\ .$$

We now make a few obvious remarks. First, if T is a flow (or cascade) then for each t,s, $T(t,s)$ is a bijection (that is, is one one and onto X). In fact, from 1.1 we have $T(t,s)\circ T(s,t)$ and $T(s,t)\circ T(t,s)$ are the identities, so $T(t,s)$ is a bijection with inverse $T(s,t)$.

Secondly suppose T is stationary (flow or cascade or any group G replacing R) and define $T^t\colon\ X \longrightarrow X$ by $T^t = T(t,o)$. Then we check that

(i) $T^{t-s} = T(t,s)$.

(ii) $T^{t+s} = T^t \circ T^s$.

(iii) T^o is the identity.

Sometimes T^t is called the flow. Observe that for a stationary cascade $T^n = (T^1)^n$.

Next, if T^t is a stationary flow on X, then X is the disjoint union of the orbits. (Same for cascades or any group.) To see this, suppose $T^{t_1} x_1 = T^{t_2} x_2$ which implies $T^t x_1 = T^{t+t_2-t_1} x_2$ so the orbit of x_1 lies in that of x_2. By symmetry they are equal, proving the assertion.

Thus in a stationary flow the orbits never intersect themselves or other orbits. For non-stationary flows this is not true as may be seen by considering a time dependent Hamiltonian system.

Before introducing more structure we mention some important examples of flows. (Some of the mathematical ideas will be further explained below.)

1.2 Examples: 1. Classical mechanical systems. A symplectic manifold (phase space) is a manifold equipped with a two form corresponding to the occurrence of variables in canonically conjugate pairs. Given this structure, each smooth function H determines in a natural way a stationary flow T^t (Hamilton's equations). This flow conserves energy ($H \circ T^t = H$) and preserves the phase volume element (Liouvilles theorem) or is measure preserving (see below). For a detailed account of these systems, see Abraham [1] and Arnold-Avez [1]. In general, smooth flows on manifolds are obtained by integrating a system of ordinary differential equations.

2. Quantum Mechanical Systems. A quantum mechanical system consists of a self adjoint operator on a Hilbert space $\mathcal{H}$. By Stone's theorem, there is a uniquely determined stationary flow T^t on $\mathcal{H}$ such that T^t is unitary for all $t \in R$. (See, for example, Yosida [1, p. 253].)

3. Heat Flow. The heat equation, $\frac{\partial f}{\partial t} = \Delta f$, determines a flow on a space of functions (or distributions) on R^n. This flow is irreversible and stationary. For example, the delta function δ is propagated only for positive times.

Notice a fundamental difference between Example 1 and 2, 3. In 1 the flow is on a set X and induces one on the functions over that set. However in 2, 3 the flow is intrinsically on a function space and does not arise from a flow on the underlying set. For example, measure preserving does not make sense for 2 and 3.

It was one of the important realizations of the 1930's that for many of the fundamental theorems (see §2 especially), only the structure of a measure space is required, and for the flow to preserve this structure.

We now recall some of the basic facts of measure theory. Good references are Halmos [2] and Berberian [1].

1.3 Definition. Let X be a set. A measure (or outer measure) on X is a map μ from the collection of all subsets of X, denoted 2^X, to the reals with ∞ adjoined such that

(i) for each $A \subset X$, $\mu(A) \geq 0$,

(ii) $\mu(\emptyset) = 0$

(iii) if $B_1, B_2, \ldots$ are subsets of X and $A \subset \bigcup_{i=1}^{\infty} B_i$ then
$$\mu(A) \leq \sum_{i=1}^{\infty} \mu(B_i) .$$

A set $A \subset X$ is called μ-measurable iff for all $B \subset X$,

$\mu(B) = \mu(A \cap B) + \mu(B \setminus A)$ where $B \setminus A = \{x: \ x \in B, x \notin A\}$.

For example, on R , Lebesgue measure is defined by

$$\mu(A) = \inf\{ \sum_{i=1}^{\infty} \ell(I_i) : \ I_i \text{ are intervals, } A \subset \bigcup_{i=1}^{\infty} I_i \}$$

where $\ell(I_i)$ is the length of I_i , and inf denotes infimum or greatest lower bound.

The basic property of measures is:

1.4 Theorem. Let μ be a measure on X and Σ denote the measurable subsets of X . Then

(i) Σ is a σ-algebra; that is if $A \in \Sigma$ then $A^c = X \setminus A \in \Sigma$ and $B_1, B_2, \ldots \in \Sigma$ implies $\bigcup_{i=1}^{\infty} B_i \in \Sigma$. (Clearly $\Sigma \neq \emptyset$, $\emptyset \in \Sigma$, $X \in \Sigma$, and $\bigcap_{i=1}^{\infty} B_i \in \Sigma$) .

(ii) if $A_1, A_2, \ldots \in \Sigma$ and $A_i \cap A_j = \emptyset$, $i \neq j$, (disjoint) then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

The proof is standard. See Halmos [2], for example. Note that if $\mu(A) = 0$ then $A \in \Sigma$. Also note that if $\{\Sigma_\alpha\}$ is any collection of σ-algebras then $\cap\{\Sigma_\alpha\}$ is also a σ-algebra.

The concept of "almost everywhere" is very useful. A proposition is said to hold a.e. (almost everywhere) iff the set of points on which it fails has measure zero. For example if $f: X \longrightarrow R$ we might say $f = 0$ a.e.

We now require that a flow on a measure space respect the measure structure. First we consider mappings in general.

1.5 Definition. A measure space is a triple (X, Σ, μ) where X is a set, μ a measure and Σ the measurable subsets. For measure spaces (X, Σ, μ), (X', Σ', μ') a map $T\colon X \longrightarrow X'$ is called measurable iff for each $A' \in \Sigma'$, $T^{-1}(A') = \{x \in X\colon T(x) \in A'\} \in \Sigma$. If, in addition, $\mu(T^{-1}(A')) = \mu(A')$ we say T is measure preserving. We say T is an isomorphism iff T is a bijection and T and T^{-1} are measure preserving. If $X = X'$ we call an isomorphism an automorphism.

Of course if T is a bijection and T, T^{-1} are measurable, T is measure preserving iff T^{-1} is.

1.6 Definition. Let X and X' be sets with measures μ and μ' respectively. Then the product measure $\mu \times \mu'$ on $X \times X' = \{(x,x')\colon x \in X \text{ and } x' \in X'\}$ is defined by

$$\mu \times \mu'(A) =$$

$$\inf\{\sum_{i=1}^{\infty} \mu(A_i)\mu'(A_i')\colon A_i \subset X,\ A_i' \subset X',\ A \subset \bigcup_{i=1}^{\infty} A_i \times A_i'\} .$$

For example the product measure on Euclidean n-space $R^n = R \times R \times \ldots \times R$ is again called Lebesgue measure. Consult a book on measure theory for the basic properties of product measure.

A measurable flow is defined as follows. (Note abuse of language.)

1.7 Definition. Let (X, Σ, μ) be a measure space and $T\colon R^2 \times X \longrightarrow X$ be a flow. We say that the flow is measurable iff

T is a measurable mapping, using Lebesgue measure on R^2 and the product measure on $R^2 \times X$. If, in addition, $T(t,s)$ is measure preserving for each $t, s \in R$ we say the flow is <u>measure preserving</u>.

Since $T(t,s)^{-1} = T(s,t)$, measure preserving means $T(t,s)$ is an automorphism.

If T is a stationary flow with $T(t-s, x) = T(t,s,x)$ then

(i) T is measurable iff T is measurable

and

(ii) the flow is measure preserving iff T_t is measure preserving for all $t \in R$.

For the proof, we have that $T = T \circ p$ where $p(t,x) = (t,o,x)$ and the composition of measurable maps is measurable. Similarly $T = T \circ s$ where $s(t,s,x) = (t-s, x)$. Finally (ii) is obvious.

We leave it to the reader to make the necessary changes for cascades, irreversible flows, etc.

Next we consider the extra topological structure which we might wish to impose:

1.8 <u>Definition</u>. A <u>topological space</u> is a pair $(X, \mathcal{T})$ where X is a set and $\mathcal{T}$ is a collection of subsets, called <u>open sets</u> such that

(i) $\emptyset, X \in \mathcal{T}$

(ii) if $A, B \in \mathcal{T}$ then $A \cap B \in \mathcal{T}$

(iii) if $\{A_\alpha\}$ is any family of open sets then $\cup\{A_\alpha\} \in \mathcal{T}$.

Often a topology arises from a metric d on X ; that is a mapping $d: X \times X \longrightarrow R$ such that

$$\text{(i)}\quad d(x,y) \geq 0 \text{ and } d(x,y) = 0 \text{ iff } x = y .$$

$$\text{(ii)}\quad d(x,y) = d(y,x) .$$

$$\text{(iii)}\quad d(x,y) \leq d(x,z) + d(z,y) .$$

The open sets are then defined by: $A \in \mathcal{J}$ iff for each $x \in A$ there is an $\varepsilon > 0$ such that $\{y \in X: d(x,y) < \varepsilon\} \subset A$. This is used on R^n , for example.

If $(X, \mathcal{J})$ is a topological space we say $A \subset X$ is closed iff $X \backslash A = A^c \in \mathcal{J}$. A subset $C \subset X$ is called compact iff for every family $\{U_\alpha\}$ of open sets with $C \subset \cup\{U_\alpha\}$, there is a finite number of the U_α , say $U_1, \ldots, U_n$ such that $C \subset \bigcup_{i=1}^{n} U_i$ (cover C) . The topology is called Hausdorff iff for each $x, y \in X$, $x \neq y$, there are disjoint open sets U, V with $x \in U$, $y \in V$ (neighborhoods of X) . We shall assume Hausdorff unless otherwise stated.

Note that closed sets obey the axioms dual to those for open sets. If $C \subset X$ is compact then it is closed. Also, in a metric space, $C \subset X$ is compact iff for each sequence $\{x_i\} \subset C$ there is a convergent subsequence. ($x_n \longrightarrow x$ iff for all neighborhoods U of x , there is an N so $n \geq N$ implies $x_n \in U$) . In R^n , a set is compact iff it is closed and bounded.

Let $(X, \mathcal{J})$ and $(X', \mathcal{J}')$ be topological spaces and $T: X \longrightarrow X'$ a map. We call T continuous iff for each $U \in \mathcal{J}'$, $T^{-1}(U) \in \mathcal{J}$. (In the case of metric spaces this is equivalent to the usual ε, δ definition).

The Borel sets on a topological space is the smallest σ-algebra containing $\mathcal{J}$. That is,

$$\mathcal{B} = \cap \{\Sigma : \mathcal{J} \subset \Sigma , \ \Sigma \text{ is a } \sigma\text{-algebra}\} .$$

(Clearly $\mathcal{J}$ may be replaced by $\mathcal{C}$, the closed sets).

For two topological spaces $(X, \mathcal{J})$ and $(X', \mathcal{J}')$ the product space $(X \times X', \mathcal{S})$ is defined as the set $X \times X'$ with topology $\mathcal{S} = \{A \subset X \times X' :$ for each $(x,y) \in A$ there is $U \in \mathcal{J}$, $V \in \mathcal{J}'$ so $x \in U$, $y \in V$ and $U \times V \subset A\}$.

1.9 Definition. A topological measure space $(X, \Sigma, \mu, \mathcal{J})$ consists of a measure space (X, Σ, μ) and a topology $\mathcal{J}$ on X such that $\mathcal{J} \subset \Sigma$ (and hence the Borel sets $\mathcal{B}$ are measurable), if $C \subset X$ is compact then $\mu(C) < \infty$ and if $U \neq \emptyset$ is open, $\mu(U) > 0$. (Often one also requires other properties of μ such as regularity.)

A continuous flow on a topological (measure) space is a flow $T: R^2 \times X \longrightarrow X$ which is continuous using the product topology. Similarly, T is a continuous measure preserving flow if in addition T is measure preserving (see 1.7).

As before, if T is stationary it is enough to check $T: R \times X \longrightarrow X$ is continuous. Cascades are similar, using the discrete topology $(Z, 2^Z)$.

The third level of specialization is to replace continuous by smooth. We now briefly sketch the main ideas. For details, see Abraham [1].

1.10 Definition. Let X be a set. An atlas on X is a collection $\{(U_\alpha, \varphi_\alpha)\}$ where $U_\alpha \subset X$, $\cup\{U_\alpha\} = X$, $\varphi_\alpha\colon\ U_\alpha \longrightarrow V_\alpha \subset R^n$ is a bijection onto an open set V_α with the property that if $U_\alpha \cap U_\beta \neq \emptyset$, the map $\varphi_\beta \circ \varphi_\alpha^{-1}$, restricted to $\varphi_\alpha(U_\alpha \cap U_\beta)$ is smooth. (All derivatives exist in R^n). Two atlases $\mathcal{A}_1$, $\mathcal{A}_2$ are equivalent iff $\mathcal{A}_1 \cup \mathcal{A}_2$ is an atlas. An equivalence class of atlases is called a differentiable structure $\mathcal{S}$ on X, and X becomes a differentiable manifold. Elements (U,φ) of $\mathcal{S}$ are called local charts.

For a manifold X we are given free a topology on X. Namely $A \subset X$ is open iff for each $a \in A$ there is a chart (U,φ) with $a \in U \subset A$.

Quite often we are also given free the structure of a measure space. For example for a mechanical system (a classical Hamiltonian system) there is given a canonical measure μ, called the phase volume. Then X is a topological measure space. In general $\mu(X) = \infty$.

If X and X' are differentiable manifolds, a mapping $T\colon\ X \longrightarrow X'$ is smooth iff for every chart (U',φ') on X' there is one on X, (U,φ) so $T(U) \subset U'$ and $\varphi' \circ T \circ \varphi^{-1}$ is smooth.

Product manifolds are defined analogously to the product of topological spaces.

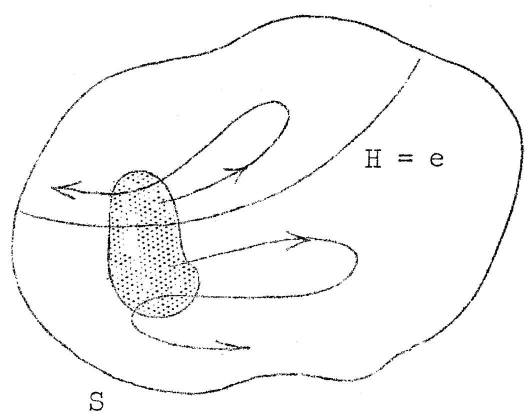
1.11 Definition. Let X be a differentiable manifold. Then a smooth flow on X is a flow $T\colon\ R^2 \times X \longrightarrow X$ which is a smooth mapping of manifolds. Similarly we may define a smooth measure preserving flow.

For example if we are given a smooth Hamiltonian on the phase space of a mechanical system we get a smooth measure preserving flow.

We now return to the general measure space setting. To illustrate the strong consequences of measure preserving, we state the following:

1.12 Poincaré's recurrence theorem. Let (X, Σ, μ) be a finite measure space $(\mu(X) < \infty)$ and $T: X \longrightarrow X$ a measure preserving transformation (1.5). Then for each $S \in \Sigma$, the following property holds for almost all $x \in S$; $T^n x \in S$ for infinitely many integers n.

This holds for bounded mechanical systems for example. Note that each point need recur only in S and not in other sets. For example, for a classical system energy surfaces partition the space.



$$(T^t)^n = T^{nt}$$

Translation on R shows that 1.12 is false if $\mu(X) = \infty$. The proof of 1.12 is simple. See Halmos [1, p. 10].

The founders of statistical mechanics argued that constants of the motion define surfaces on which the motion takes place and that in the remaining degrees of freedom the motion should be "ergodic" or at least "quasi-ergodic" the first meaning that there is an orbit (or positive orbit) filling the entire space and the second that every orbit is dense. The situation is, in fact much more complicated. It is possible to have an orbit filling the space (a flow with one orbit) although this a priori excludes continuous flows. The more modern definition of ergodic is given below.

Recall that if $(X, \mathcal{J})$ is a topological space and $A \subset X$, the <u>closure</u> of A is defined by

$$\overline{A} = \cap \{C \subset X: \ C \text{ is closed and } A \subset C\}$$

and is a closed set. A is called <u>dense</u> iff $\overline{A} = X$.

1.13 <u>Definition</u>. Let $(X, \mathcal{J})$ be a topological space and $T: R^2 \times X \longrightarrow X$ a flow. We say T is <u>topologically transitive</u> iff T has a dense positive orbit. If, in addition X is a measure space and almost every positive orbit is dense we say T is <u>minimal</u>.

Let (X, Σ, μ) be a measure space and $T: R^2 \times X \longrightarrow X$ a measurable flow. We say T is <u>ergodic</u> (or <u>irreducible</u> or <u>metrically transitive</u>) iff $A \in \Sigma$, $T(t,s)^{-1}(A) = A$ for all $t \geq s$ implies either $\mu(A) = 0$ or $\mu(A^c) = 0$.

Let (X, Σ, μ) be a finite measure space and $T: R \times X \longrightarrow X$ a stationary flow. We say T is <u>mixing</u> (or <u>strongly mixing</u>) iff for all $A, B \in \Sigma$,

$$\operatorname*{limit}_{t \to +\infty} [\mu(T^t(A) \cap B) - \mu(A)\ \mu(B)/\mu(X)] = 0 .$$

Also, T is called <u>weakly mixing</u> iff for all $A, B \in \Sigma$,

$$\operatorname*{limit}_{t \to +\infty} \frac{1}{t} \int_0^t |\mu(T^s(A) \cap B) - \mu(A)\ \mu(B)/\mu(X)| ds = 0 .$$

(Similar definitions hold for cascades.)

Recall now the definition of integral on a measure space (X, Σ, μ).

Here measurability is essential. For $f: X \longrightarrow R$ measurable, define $f^+(x) = \max(f(x), 0)$ and $f^- = (-f)^+$. A step function is a measurable function assuming only a finite number of values. There are step functions so $f_n^+ \uparrow f^+$ pointwise and $f_n^- \uparrow f^-$ pointwise. $\int f_n^+ \, d\mu = \Sigma f_n^+(A_j)\, \mu(A_j)$ where A_j are the sets on which f_n^+ is constant. We define $\int f d\mu = \lim\limits_{n\to\infty} \int f_n^+ \, d\mu - \lim\limits_{n\to\infty} \int f_n^- \, d\mu$ if both numbers are not $+\infty$, and call f <u>integrable</u> if $\int f d\mu$ is finite.

In particular, on Z with the discrete measure the integral is just summation.

One of the most useful theorems is:

1.14 <u>Lebesgue's dominated convergence theorem</u>. Suppose $f_n, f: X \longrightarrow R$ are measurable on a measure space X and there is an integrable function g so $|f_n(x)| \leq |g(x)|$ for all n, and almost all x. Then if $f_n(x) \longrightarrow f(x)$ for almost all x,

$$\lim_{n\to\infty} \int f_n \, d\mu = \int \lim_{n\to\infty} f_n \, d\mu = \int f d\mu .$$

We now return to 1.13 and discuss the basic relationship between the notions defined.

First, mixing implies weak mixing, for given $\varepsilon > 0$ choose T^ε so $t \geq t_\varepsilon$ implies $|\mu(T^t(A) \cap B) - \mu(A)\, \mu(B)/\mu(X)| < \varepsilon/2$. Choose T_0 so $T \geq T_0$ implies $\frac{1}{T} \int_0^{t_\varepsilon} |\mu(T^t(A) \cap B) - \mu(A)\, \mu(B)/\mu(X)| dt \leq \varepsilon/2$. Then we have

$$\frac{1}{T}\int_0^T |\mu(T^t(A) \cap B) - \mu(A)\,\mu(B)/\mu(X)|\,dt$$

$$\leq \frac{1}{T}\int_0^{t_\varepsilon} |\mu(T^t(A) \cap B) - \mu(A)\,\mu(B)/\mu(X)|\,dt$$

$$+ \frac{1}{T}\int_{t_\varepsilon}^T |\mu(T^t(A) \cap B) - \mu(A)\,\mu(B)/\mu(X)|\,dt < \varepsilon .$$

Secondly, weak mixing implies ergodicity, for if $\mu(X\setminus A) \neq \emptyset$, and $T^{-t}(A) = A$, choose $B = X\setminus A$, so that

$$\frac{1}{T}\int_0^T |\mu(T^t(A) \cap B) - \mu(A)\,\mu(B)/\mu(X)|\,dt$$

$$= \mu(A)\,\mu(B)/\mu(X)$$

since $T^t(A) \cap B = \emptyset$. Hence $\mu(A) = 0$.

Next, if T is an ergodic stationary flow [resp. cascade] on a topological measure space which is second countable (there is a countable collection of open sets $\{U_n\}$ such that every open set contains some U_n) , then T is minimal.

For the proof, $x \in X$ has non dense positive orbit iff there is a U_n so $x \in A_n = \bigcap_{t \geq 0} X\setminus T^t(U_n)$. Now $T^{-t}(A_n) = A_n$ and $A_n^c \supset U_n$, $\mu(U_n) \neq 0$ so $\mu(A_n) = 0$ by ergodicity. Thus all points with non dense orbits lie in $\cup A_n$ which has measure zero (1.4 (ii)). For further discussion along these lines for cascades, see Halmos [1, pp. 25-30].

mixing $\longrightarrow$ weak mixing $\longrightarrow$ ergodicity $\longrightarrow$ minimality $\longrightarrow$

topologically transitive.

The name irreducibility arises as T is not ergodic iff there are two disjoint measurable sets $A, B \in \Sigma$ with $A \cup B = X$ and both invariant under the flow (obvious).

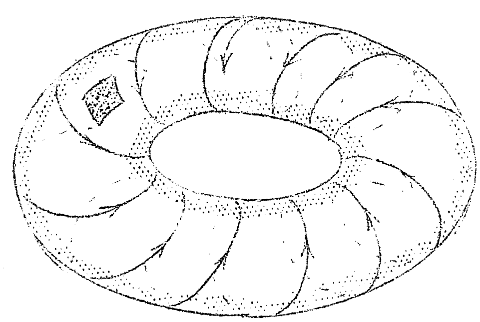
In general, ergodicity is not at all obvious, but seems an essential property for statistical mechanics.

Classical Hamiltonian systems are never ergodic. In fact for a given energy e, the two sets $\{x: H(x) \geq e\}$ and $\{x: H(x) \leq e\}$ are invariant with positive measure in general (unless H is constant, whence the flow is stationary and every set is invariant).

Thus we have a chance for ergodicity only on the <u>energy surfaces</u> $H^{-1}(e)$. In fact it is known that for almost all e, $H^{-1}(e)$ is a manifold (Sard's theorem) on which the flow is defined, and that this manifold possesses a measure invariant under the flow. For the proof, see Abraham [1, §15].

For the intuitive difference between ergodic and mixing we consider two examples (Arnold-Avez [1, pp. 4, 8].) Note that, roughly speaking ergodic means that every mass visits every other mass, while mixing means that after long times any mass becomes uniformly spread over X.

The first example is group translation on the torus T^2 through an irrational slope. (Explicitly, $T^t(x) = x + tc$ where c is irrational on the covering space.) The flow is smooth measure preserving, ergodic but not mixing.

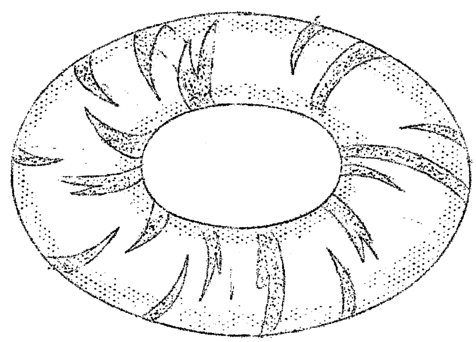


Note that a small region retains its identity under the flow.

The second example is represented on R^2 by

$$T^t(x,y) = (x+ty,\ tx + (1+t^2)y)$$

and is a measure preserving mixing flow. After long times $t \geq 0$, a small mass is shredded and spread uniformly. (See Arnold Avez [1, p. 8] for details.)



It is the property of mixing, or stronger a K-system (see below) which is responsible for the "irreversibility" phenomena in finite systems.

§2. Koopmanism

Koopman's idea was to transfer the flow on a measure space to a flow on a space of functions and to relate spectral properties of the new flow with ergodicity properties of the original flow.

To do this, we first set up some familiar machinery.

2.1 Definition. Let (X, Σ, μ) be a measure space and $p \in R$, $1 \leq p < \infty$. Let

$$L^p(X,\mu) = \{f\colon\ X \longrightarrow R : |f|^p \text{ is integrable}\}$$

and $\|f\|_p = (\int |f|^p \, d\mu)^{1/p} < \infty$.

For measure spaces (X, Σ, μ) and (X', Σ', μ'), and $T\colon X \longrightarrow X'$ measurable, define the measure $T_*\mu$ on X' by

$$T_*\mu(A) = \mu(T^{-1}(A))$$

for $A \in 2^X$.

In this definition we may replace R by $\mathcal{C}$, the complex numbers. The integral extends linearly to maps $f\colon X \longrightarrow \mathcal{C}$.

The basic change of variables theorem is as follows.

2.2 Theorem. Suppose $T\colon X \longrightarrow X'$ is measurable and $f \in L^1(X', T_*\mu)$. Then $f \circ T \in L^1(X,\mu)$ and

$$\int f d(T_*\mu) = \int (f \circ T)\, d\mu .$$

The proof is a simple exercise. See Halmos [2, p. 163]. The next theorem is not so simple, but is standard.

2.3 Theorem. If (X, Σ, μ) is a measure space and $1 \leq p < \infty$ then $L^p(X,\mu)$ is a Banach space. That is, $L^p(X,\mu)$ is a real (or complex) vector space and $\| \|_p$ satisfies:

(i) $\|cf\|_p = c\|f\|_p$ for $c \in R$ (or $\mathbb{C}$).

(ii) $\|f+g\|_p \leq \|f\|_p + \|g\|_p$ (Minkowski's inequality).

(iii) $\|f\|_p = 0$ iff $f = 0$ a.e. and as a metric space with $d(f,g) = \|f-g\|_p$, $L^p(X,\mu)$ is complete; that is if $f_n \in L^p(X,\mu)$ is a Cauchy sequence (for all $\varepsilon > 0$ there is an integer N so $n, m \geq N$ implies $d(f_n,f_m) < \varepsilon$) then there is an $f \in L^p(X,\mu)$ so $f_n \longrightarrow f$.

In particular, $L^2(X,\mu)$ is a Hilbert space; that is, there is a bilinear map $<,> : L^2(X,\mu) \longrightarrow R$ so $\|f\|_2^2 = <f,f>$. In fact, $<f,g> = \int fg\, d\mu$ or, in the complex case, $<f,g> = \int \bar{f}g\, d\mu$ where $\bar{}$ denotes complex conjugation.

A flow on a measure space induces one on $L^p(X,\mu)$ in a natural way.

2.4 Definition. Let (X, Σ, μ) be a measure space and $T: R^2 \times X \longrightarrow X$ a measure preserving flow on X. Then for $1 \leq p < \infty$, define

$$U: R^2 \times L^p(X,\mu) \longrightarrow L^p(X,\mu)$$

by

$$[U(t,s)f](x) = f(T(s,t)x)$$

or $U(t,s)f = f \circ T(t,s)^{-1}$.

The map U is called the <u>propagator</u> of the flow T.

The basic properties of U are easily deduced from 2.2 as follows.

2.5 <u>Theorem.</u> In 2.4 we have

(i) U is well defined.

(ii) U is a flow on $L^p(X,\mu)$ and is stationary if T is a stationary flow.

(iii) for each $(s,t) \in R^2$, $U(s,t)$ is an isomorphism. That is, $U(s,t)$ [and $U(s,t)^{-1} = U(t,s)$ by (ii)) satisfies $\|U(s,t)f\|_p = \|f\|_p$ for each $f \in L^p(X,\mu)$. In particular, if $p = 2$, $U(s,t)$ is unitary, or $\langle U(s,t)f, \ U(s,t)g\rangle = \langle f,g\rangle$.

<u>Proof.</u> As T is measure preserving, $T(t,s)_*\mu = \mu$ so that if $|f|^p \in L^1(X,\mu)$, then $|f|^p \circ T(t,s)^{-1} = |f\circ T(t,s)^{-1}|^p \in L^1(X,\mu)$ by 2.2, so (i) holds. Also from 2.2, $\int |f|^p \, d\mu = \int |f\circ T^{-1}(t,s)|^p \, d\mu$ so that (iii) is clear. Finally, (ii) follows at once from the definitions. □

The following is sometimes a convenient criterion for ergodicity.

2.6 <u>Theorem.</u> Let T be a measurable flow on a measure space (X, Σ, μ). Then T is ergodic iff for each $f\colon X \longrightarrow R$ (or $\mathbb{C}$) measurable, with $f = f\circ T(t,s)$ a.e. for all $t, s \in R$ (i.e.: f is a <u>constant of the motion</u>) implies f is constant a.e.

<u>Proof.</u> If T is not ergodic there are two sets A, B invariant under T and $\mu(A) \neq 0$, $\mu(B) \neq 0$ and $A \cap B = \emptyset$, $A \cup B = X$. Obviously

$f = 1$ on A, 0 on B is a constant of the motion. Conversely, suppose T is ergodic and f is a constant of the motion. If f is not constant there is a number $a \in R$ so $A = \{x: f \leq a\}$, $B = \{x: f > a\}$ have positive measure. Clearly A, B are invariant, implying T is not ergodic. $\square$

For further discussion along these lines see Halmos [1, p. 25], Wightman [1] and Abraham [1, §16, 29].

Recall that a measure space is called <u>σ-finite</u> iff there are $S_i \in \Sigma$, $i = 1,2,3,\ldots$ such that $X = \cup S_i$ and $\mu(S_i) < \infty$. (This condition is almost always satisfied in practice.)

As a Hilbert space is a topological space, it makes sense to talk about continuous flows $U: R^2 \times \mathcal{H} \longrightarrow \mathcal{H}$. (See 1.9). In the case of a unitary flow, as in 2.5 this is equivalent to continuity of either of the following maps

$$\text{(i)} \quad (t,s) \longmapsto \langle f, U(t,s)g\rangle \quad \text{for each } f, g \in \mathcal{H}.$$

$$\text{(ii)} \quad (t,s) \longmapsto \langle f, U(t,s)f\rangle \quad \text{for } f \in \mathcal{H}.$$

(More generally when $U(t,s)$ is a bounded linear transformation.)

2.7 <u>Theorem</u> (<u>Koopman</u>). Let T be a measure preserving flow on a σ-finite measure space (X, Σ, μ). Then the propagator U of T defines a continuous flow on $L^2(X,\mu)$.

For the proof see Wightman [1, lecture 7] and Dunford-Schwartz [1, pt I, p. 616].

About the time Koopman proved this theorem, the analysis in Hilbert spaces was being strongly advanced (1930's). In particular, Stone's theorem (see Yosida [1]) give the following corollary of 2.7.

2.8 Theorem. In 2.7, suppose T is stationary. Then there is a unique self adjoint operator H on $L^2(X,\mu)$ such that iH is the infinitesimal generator of U and we write $U_t = \exp(iHt)$. (That $L^2(X,\mu)$ be the complex L^2 space is essential.)

To explain what this means, we recall a few definitions:

2.9 Definitions. Let $\mathcal{H}$ be a Hilbert space, $D(T) \subset \mathcal{H}$ and $T: D(T) \longrightarrow \mathcal{H}$ a linear map. We say T is self adjoint iff

(i) $D(T)$ is dense in $\mathcal{H}$ (1.8).

(ii) for $f, g \in D(T)$, $\langle Tf,g\rangle = \langle f,Tg\rangle$.

(iii) there are no $g \in \mathcal{H}\setminus D(T)$ with the property that there is a $g_1 \in \mathcal{H}$ so $\langle Tf,g\rangle = \langle f,g_1\rangle$ for all $f \in D(T)$. (Maximality of domain.)

If $U: R \times \mathcal{H} \longrightarrow \mathcal{H}$ is a continuous stationary flow on $\mathcal{H}$, the infinitesimal generator of U, say A is defined on the set

$$D(A) = \{f \in \mathcal{H} : \lim_{t \to 0} (U^t f - f)/t \text{ exists}\}$$

and on $D(A)$, A is defined as this limit. (The derivative with respect to t).

In case the flow T arises from a smooth Hamiltonian system we can say exactly what iH of 2.8 is, in terms of the original Hamiltonian.

It is the (generalized) Lie derivative iL_{X_H} on

$$\{f \in \mathcal{H} : L_{X_H} f \in \mathcal{H}\} = D(L_{X_H})$$

where X_H is the operator canonically associated with the classical Hamiltonian H. For details, see Abraham [1, Chapter III].

To be able to understand Koopman's spectral condition, we state one version of the spectral theorems. For details see Yosida [1].

2.10 Spectral Theorem for Self Adjoint Operators. Let $\mathcal{H}$ be a Hilbert space and $A: D(A) \longrightarrow \mathcal{H}$ a self-adjoint operator on $\mathcal{H}$. Suppose for some $f \in D(A)$, $A^n f \in D(A)$ and $\mathcal{H}_0$ denotes the space spanned by $(f, Af, A^2 f, \ldots)$. Then there is a measure σ on R such that $\mathcal{H}_0$ is isomorphic to $L^2(R,\sigma)$ and if φ denotes this isomorphism,

$$\varphi(Af)(x) = xf(x) ,$$

or A corresponds to multiplication by x. Here, σ is called the spectrum of A, or spectral measure.

In general, A need not have any (non zero) eigenfunctions in $\mathcal{H}_0$. Eigenvalues correspond to singular, isolated points of the spectrum σ.

(One can replace $\mathcal{H}_0$ by $\mathcal{H}$ if $\mathcal{H}$ is decomposed into a direct integral).

Similarly, for unitary operators we have: (Since U is uniformly continuous if U is unitary, its domain may be taken all of $\mathcal{H}$).

2.11 Spectral Theorem for Unitary Operators. Let U be a unitary operator on a Hilbert space $\mathcal{H}$ and for some $f \in \mathcal{H}$ let $\mathcal{H}_0$ be generated by $\{f, Uf, U^{-1}f, U^2f, \ldots\}$. Then there is a measure σ on the unit circle S^1 and an isomorphism $\varphi\colon \mathcal{H}_0 \longrightarrow L^2(S^1, \sigma)$ such that $\varphi(Ug)(\lambda) = \lambda g(\lambda)$. Again, σ is called the spectrum of U.

To recover completeness theorems (as is often "postulated" in elementary quantum mechanics) one must pass to generalized functions (distributions).

Koopman's connection between the spectrum of the propagator U_t, (or the generator iH) is given as follows. (An eigenvalue is simple iff the space of eigenfunctions corresponding is one dimensional).

2.12 Theorem. Let (X, Σ, μ) be a finite measure space, T a measure preserving stationary flow on X and U its propagator. Then

(i) T is ergodic iff 1 is a simple eigenvalue of U_t for all $t \in R$.

If (i) holds and $0 \leq \theta < 2\pi$, and M_θ is

$$M_\theta = \{f \in L^2(X,\mu)\colon \; U_t f = e^{i\theta t} f$$
$$(\text{or } \; Hf = \theta f)\}$$

then either $M_\theta = \{0\}$ or M_θ is one dimensional.

(ii) T is weakly mixing iff 1 is a simple eigenvalue of U_t and there are no other eigenvalues. (See 2.11).

For the proof, see Wightman [1, lecture 16].

It can be shown (using distribution theory) that if T arises from a smooth (Hamiltonian) flow then T is uniquely determined by the spectrum of the propagator U_t (or H) .

In general, spectral invariants are not enough to get properties of the flow, such as the Kolmogorov-Sinai entropy. (See §5).

§3 Ergodic Theorems

In this section we state the ergodic theorems of Von Neumann and Birkhoff. These theorems deal with time and phase averages of functions and their equality under ergodicity.

First, recall that if $\mathcal{H}$ is a Hilbert space and $C \subset \mathcal{H}$ is a closed subspace every $x \in \mathcal{H}$ can be uniquely written $x = x_1 + x_2$ where $x_1 \in C$ and $\langle x_2, y\rangle = 0$ for all $y \in C$. The map $x \longrightarrow x_1$ is called the projection onto C.

We begin with Von Neumann's theorem dealing with mean convergence.

3.1. Theorem (Mean Ergodic Theorem). Let $\mathcal{H}$ be a Hilbert space and U^t a stationary unitary flow on $\mathcal{H}$. Let $\mathcal{H}_o = \{f \in \mathcal{H}: U^t f = f$ for all $t \in R\}$ and $P: \mathcal{H} \longrightarrow \mathcal{H}_o$ the projection onto $\mathcal{H}_o$. Then for each $f \in \mathcal{H}$ we have

$$\operatorname*{limit}_{t \to \infty} \frac{1}{t} \int_0^t U^s f \, ds = Pf$$

(using the topology of $\mathcal{H}$.)

In case $\mathcal{H} = L^2(X,\mu)$, we define

$$\left(\int_0^t U^s f \, ds \right)(x) = \int_0^t U^s f(x) \, ds$$

In general the integral is defined similarly to the Lebesque integral, by step function approximation. For details, see Yosida [1, p. 132].

For the proof of 3.1 in the discrete case, see Halmos [1, p. 13-17] and in general, see Wightman [1].

From 3.1 we may deduce the following useful corollary:

3.2. <u>Corollary.</u> Suppose (X,Σ,μ) is a finite measure space and T^t is a stationary measure preserving flow with propagator U^t (2.4). Then T is ergodic iff for each $f \in L^2(X,\mu)$ we have

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t U^t f \, dt = \left(\int f d\mu \right) / \mu(X).$$

<u>Proof.</u> If T^t is ergodic, then by 2.6, $\mathcal{H}_0$ of 3.1 is one dimensional, and writing

$$f = \left(\int f d\mu \right) / \mu(X) + \left[f - \left(\int f d\mu \right) / \mu(X) \right]$$

we see that $Pf = \left(\int f d\mu \right) / \mu(X)$. (Note that $f_0 T^t = f$ if $U^t f = f$)

Conversely, if Pf is $\left(\int f d\mu \right) / \mu(X)$ then $U^t f = f$ implies f is constant. Hence by the proof of 2.6, T^t is ergodic. □

In the case of a classical system, f square integrable on the phase space (or energy surface) is often called a classical "observable".

Next we consider Birkhoff's theorem dealing with convergence almost everywhere.

3.3 Theorem (Individual Ergodic Theorem). Let T^t be a stationary measure preserving flow on a measure space (X,Σ,μ) and $f \in L^1(X,\mu)$. Then

$$\hat{f}(x) = \lim_{t \to \infty} \frac{1}{t} \int_0^t T^s f(x)\, ds$$

exists for almost all $x \in X$, $\hat{f} = \hat{f} \circ T^t$ and $\hat{f} \in L^1(X,\mu)$. If X is a finite measure space, we have further,

$$\int f d\mu = \int \hat{f} d\mu$$

Again, see Halmos [1, p. 18-21] and Wightman [1] for the proof.

3.4. Corollary. If, in 3.3, $\mu(X) < \infty$, then T^t is ergodic iff for all $f \in L^1(X,\mu)$, $\hat{f}$ is constant almost everywhere. If this is the case, then

$$\hat{f}(x) = \int f d\mu \ / \ \mu(x) \qquad \text{a.e.}$$

Proof. If T^t is ergodic then $\hat{f}$ is constant almost everywhere by 2.6. Conversely if $f \in L^2(X,\mu)$ has $f = f \circ T^t$ then $\hat{f} = f$, so $\hat{f}$ constant implies T^t is ergodic. The last part is clear from 3.3. □

In contrast to 3.1, 3.3 gives convergence in $L^2(X,\mu)$.

3.5. Corollary. In 3.3, $\frac{1}{t} \int_0^t U^s f\, ds$ converges to $\hat{f}$ in $L^1(X,\mu)$.

Proof. We may suppose $f \geq 0$. Let $f_n = \sup\{f, U^1 f, \cdots, U^n f\}$ and $g = \lim f_n$. Now $f_n \nearrow g$ and $\int f_n d\mu = \int f d\mu$. Hence g is integrable. In 3.3 we have $\frac{1}{n} \int_0^n U^s f \, ds$ converging pointwise to $\hat{f}$ and each is bounded by g, so by 1.14, it converges to $\hat{f}$ in $L^1(X,\mu)$. Since any sequence $t_n \longrightarrow \infty$ has $\frac{1}{t_n} \int_0^{t_n} U^s f \, ds$ converging, $\frac{1}{t} \int_0^t U^s f \, ds$ converges. □

If $f \in L^1(X,\mu)$ and $L^2(X,\mu)$, then $\hat{f}$ of 3.1 coincides with $\hat{f}$ of 3.3, for the above argument shows we have convergence in $L^2(X,\mu)$ as well.

The ergodic theorems above show that the problem of showing time averages equal phase averages is equivalent to showing that the flow is ergodic. In general this is not obvious. As we have seen, ergodicity means that there are no more constants of the motion. In fact, sometimes completely unsuspected constants of the motion have been found for some systems.

However, there is hope, for Sinai has anounced that the system consisting of hard spheres in a box is indeed ergodic (Sinai [2]). (The proof has yet to appear.) This theorem goes a long way toward justifying the basic ideas of statistical mechanics, even though a great deal remains to be done.

For further discussion of Sinai's theorem, see Wightman [1].

Birkoff's theorem (3.3,3.4) determines the measure μ uniquely, up to a constant as follows. Recall that a Borel measure is a measure for which open sets are measurable.

3.6. Theorem. Let M be the phase space of a Hamiltonian system, $H: M \longrightarrow R$ the energy and $H^{-1}(e)$ an energy surface (submanifold). There is at most one (regular) Borel measure μ on $H^{-1}(e)$ such that

(i) $\mu(H^{-1}(e)) = 1$,

(ii) the flow on $H^{-1}(e)$ is ergodic and measure preserving.

Furthermore, if μ has any point $x \in H^{-1}(e)$ with positive measure, then μ is concentrated at a single point.

Proof. Let $U \subset H^{-1}(e)$ be an open set and X_U its characteristic function ($X_U = 1$ on U, zero on U^c). Then by 3.4 we have

$$\mu(U) = \int X_U \, d\mu = \lim_{t \to \infty} \frac{1}{t} \int_0^t U^s f \, dt$$

so that μ is uniquely determined.

For the last part, suppose $m \in M$ and $\mu(m) > 0$. Let m_t be the orbit through m. As T^t is measure preserving $\mu(m_t) = \mu(m) > 0$. But by (i) this means $m_t = m$. Then m is an invariant set, so by ergodicity, $\mu(X \setminus \{m\}) = 0$. □

If there is a measure μ satisfying these condition, it is called the microcanonical ensemble.

§4 Statictical Regularity

Here we discuss statistical regularity introduced by Hopf [1]. It is important for the intuition it provides about mixing, in terms of probability. It is the key to understanding how deterministic classical laws (i.e. a flow) lead to non-deterministic, that is, probabilistic laws.

To see how probability enters consider on ergodic flow T^t on a finite measure space (X,Σ,μ). We interpret, for $B\epsilon\Sigma$, $\mu(B)/\mu(X)$ as the probability of the event B; that is, the probability that a point $x\epsilon X$ will lie in B. This is the usual formulation of probability theory. On the other hand, it is reasonable since $\frac{1}{t}\int_o^t X_B(x_t)dt$ converges to $\mu(B)/\mu(X)$ (3.4) and represents the fraction of time that a point x lies in B.

In other words, the probability that x lies in B equals the fraction of time x spends in B.

As we observed in 3.6, this property forces us to adopt only one measure μ for a probabilistic interpretation (microcanonical ensemble).

It was in terms of a probabilistic description of this kind that Boltzman reinterpreted his H-theorem after the attacks of Poincaré, Loschmit and Zermelo on his conceptual foundations. At least, this seems to be the correct interpretation according to the decoding of the Ehrenfest's in their highly recommended little book (Ehrenfest P. and Ehrenfest T. [1]).

The above description indicates the physical significance of ergodicity. For mixing we owe the interpretation to Gibbs [1, p. 144] for his water-dye analogy but even more to Hopf [1] for his sucessful explanation of the 'roulette wheel problem'. That is, how can a deterministic system give rist to probabilistic laws?

The main idea is that in playing roulette the initial conditions are properly described by a probability distribution rather than a specific point. For $f \in L^1(X,\mu)$ with $\int f d\mu = 1$, and $A \subset X$, the probability of the event A after a time t is

$$\int_A U^t f d\mu = \int_A f(x_{-t})\, d\mu(x) = \int f(x)\, X_A(x_t) d\mu(x)$$

(by 2.2). We think of the probability distribution f evolving in time according to U^t.

Our intuition tells us that if the process is random, the probability of obtaining the event A should depend only on the system and not on the initial distribution. This is exactly the definition adopted by Hopf.

4.1 <u>Definition.</u> Let T^t be a stationary measure preserving flow on a measure space (X,Σ,μ). An (event) $A \in \Sigma$ is called <u>statistically regular</u> if there is a number $P(A)$ such that for every $f \in L^1(X,\mu)$,

$$\lim_{t \to \infty} \left(\int_A U^t f d\mu\right) / \left(\int f d\mu\right) = P(A)$$

Then we have:

4.2 <u>Theorem (Hopf).</u> Let T^t be a stationary measure preserving flow on a finite measure space. Then T^t is mixing (1.13) iff every measurable set is statistically regular. In this case $P(A) = \mu(A)/\mu(X)$ (taking $f = 1$).

Proof. If $A \in \Sigma$ is statistically regular and $B \in \Sigma$, let $f = X_B$ so that

$$\int_A U^t f \, d\mu = \mu(T^t A \cap B)$$

and

$$\underset{t \to \infty}{\text{limit}}\ \mu(T^t A \cap B) = \mu(B)\mu(A)/\mu(X).$$

Hence T^t is mixing.

Conversely suppose T^t is mixing. For $A \in \Sigma$ we must show A is statistically regular. The condition holds for $f = X_B$ and by addition for any step function. Suppose $f \geq 0$, $f \in L^1(X,\mu)$ and $f_n \nearrow f$ where f_n is a step function. Now $\underset{n \to \infty}{\text{limit}} \int_A U^t f_n d\mu = \int_A U^t f_n d\mu$ and the limit is uniform in t by 1.14. Hence

$$\begin{aligned}
\underset{t \to \infty}{\text{limit}} \int_A U^t f d\mu &= \underset{n \to \infty}{\text{limit}}\ \underset{t \to \infty}{\text{limit}} \int U^t f_n d\mu \\
&= \underset{n \to \infty}{\text{limit}}\ \frac{\mu(A)}{\mu(X)} \int f_n d\mu \\
&= \frac{\mu(A)}{\mu(X)} \int f d\mu.
\end{aligned}$$

(The interchange of limits is justified by Apostol [1, p. 394] for example.) □

For an alternative proof, see Wightman [1].

As a corollary, if a roulette wheel is mixing then there is no way possible to beat the wheel.

The work of Hopf went virtually unnoticed by physicists with the exception of N.S. Krylov [1]. He attempted to relate it to relaxation phenomena in a hard sphere gas. Rigorizing his notions (following Sinai) will probably require that the flow is at least mixing. Properties stronger than the mixing property motivated by these studies are defined in the next section.

§5. Entropy, K-Systems and C-Systems

For statistical mechanics, the notion of a K-system may be even more fundamental and far reaching than ergodicity or mixing. This section gives the definitions and basic properties (without proofs), which are largely due to Kolmogorov and Sinai.

We begin with the entropy of a partition:

5.1 Definition. A probability space is a measure space (X, Σ, μ) with $\mu(X) = 1$ (normalization). A finite partition $\mathcal{A}$ of X is a finite collection $\{A_1,\dots,A_n\}$ of measurable sets which are disjoint and cover X. ($A_i \cap A_j = \emptyset$ if $i \neq j$ and $\bigcup_{i=1}^{n} A_i = X$). The entropy of $\mathcal{A}$ is defined as the real number

$$H(\mathcal{A}) = -\sum_{i=1}^{n} \mu(A_i)\, \ell n\, \mu(A_i)$$

(ℓn meaning logarithm) with $0\, \ell n\, 0 = 0$.

A countable partition with its entropy is defined similarly.

For two partitions $\mathcal{A}$, $\mathcal{B}$ the conditional entropy is defined by

$$H(\mathcal{A} \mid \mathcal{B}) = \Sigma\, \{\mu(A \cap B)\, \ell n\, \mu(A \cap B)/\mu(B) : \; A \in \mathcal{A}, B \in \mathcal{B}\}$$

so that if $\mathcal{B} = \{X\}$, $H(\mathcal{A} \mid \mathcal{B}) = H(\mathcal{A})$.

Notice that $0 \leq H(\mathcal{A}) \leq \ell n\, n$ if $\mathcal{A}$ has n elements. Shannon [1] introduced the concept of entropy to measure the 'information' or 'randomness' of a partition $\mathcal{A}$. (More specifically for letters of the alphabet being the partition of a message). Kolmogorov [1,2] introduced the general concept for dynamical systems with considerable improvement by Sinai [3]. Some intuition behind the definition is given in Billingsly [1].

5.2 Definition. Suppose $\mathcal{A}$ and $\mathcal{B}$ are finite partitions of a probability space X. The common refinement is defined to be the finite partition

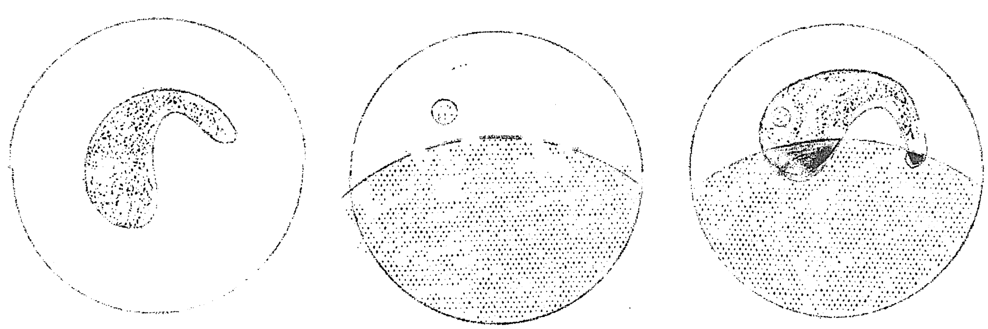
$$\mathcal{A} \vee \mathcal{B} = \{A \cap B : \; A \in \mathcal{A} \;, B \in \mathcal{B}\}$$

More generally for a countable family of finite or countable partitions $\{\mathcal{A}_n\}$ we define

$$\vee \{\mathcal{A}_n\} = \{\bigcap_n A_n : \; A_n \in \mathcal{A}_n\} .$$

If $\mathcal{A}$ and $\mathcal{B}$ are (finite) partitions of X, we say $\mathcal{A} \leq \mathcal{B}$ iff for each $B \in \mathcal{B}$ there is an $A \in \mathcal{A}$ so $B \subset A$, or $\mathcal{B}$ is a refinement of $\mathcal{A}$.

For a partition $\mathcal{A}$ let $\Sigma(\mathcal{A})$ denote the σ-algebra generated by $\mathcal{A}$ (if $\mathcal{A}$ is finite, $\Sigma(\mathcal{A})$ consists of finite unions of members of $\mathcal{A}$).



$+ \phi$

$\mathcal{A}$ $\qquad$ $\mathcal{B}$ $\qquad$ $\mathcal{A} \vee \mathcal{B}$

Clearly $\mathcal{A} \leq \mathcal{B}$ iff $\Sigma(\mathcal{A}) \subset \Sigma(\mathcal{B})$ so $\leq$ is a partial ordering and $\mathcal{A} = \mathcal{B}$ iff $\Sigma(\mathcal{A}) = \Sigma(\mathcal{B})$. Also, $\Sigma(\mathcal{A} \vee \mathcal{B}) = \Sigma(\mathcal{A} \cup \mathcal{B})$ and $\mathcal{A} \vee \mathcal{B} = \sup(\mathcal{A}, \mathcal{B})$ (that is, $\mathcal{A} \vee \mathcal{B} \geq \mathcal{A}, \mathcal{B}$ and $\mathcal{C} \geq \mathcal{A}, \mathcal{B}$ implies $\mathcal{C} \geq \mathcal{A} \vee \mathcal{B}$). (One can similarly define $\wedge$ so the (finite) partitions form a lattice. This is the reason for the notation $\vee$, $\wedge$.)

Some of the basic properties of entropy are as follows:

5.3 <u>Theorem.</u> Let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ be finite partitions of a probability space X. Then

(i) $H(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C}) = H(\mathcal{A} \mid \mathcal{C}) + H(\mathcal{B} \mid \mathcal{A} \vee \mathcal{C})$.

(ii) $H(\mathcal{B} \mid \mathcal{C}) \leq H(\mathcal{A} \mid \mathcal{C})$ if $\mathcal{B} \leq \mathcal{A}$.

(iii) $H(\mathcal{A} \mid \mathcal{C}) \leq H(\mathcal{A} \mid \mathcal{B})$ if $\mathcal{C} \geq \mathcal{B}$.

(iv) $H(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C}) \leq H(\mathcal{A} \mid \mathcal{C}) + H(\mathcal{B} \mid \mathcal{C})$ ((i), (iii) $\Rightarrow$ (iv)).

(v) if $T: X \longrightarrow X$ is measure preserving then $T^{-1}\mathcal{A} = \{T^{-1}A: A \in \mathcal{A}\}$ is a finite partition and $H(T^{-1}\mathcal{A} \mid T^{-1}\mathcal{B}) = H(\mathcal{A} \mid \mathcal{B})$.

The proof is quite easy. See Billingsly [1, p. 78].

Next we consider the entropy of a transformation:

5.4 <u>Definition.</u> Let $T: X \longrightarrow X$ be a measure preserving transformation on a probability space. Let $\mathcal{A}$ be a finite partition and

$$h(\mathcal{A}, T) = \lim_{n \to \infty} \frac{1}{n} H(\mathcal{A} \vee T^{-1}\mathcal{A} \vee \ldots \vee T^{-(n-1)}\mathcal{A})$$

and

$$h(T) = \sup\{h(\mathcal{A}, T): \mathcal{A} \text{ is a finite partition}\}$$

called the Kolmogorov-Sinai entropy of T.

From 5.3 it follows that the above limit exists and is finite (in fact the sequence is decreasing) although $h(T)$ may be infinite. See Billingsly [1, p. 81-2].

Intuitively we think of $h(\mathcal{A}, T)$ as a measure of the randomness (or information) given to $\mathcal{A}$ per application of T. In other words, it measures the mixing power of T. Taking the supremum optimizes the choice of $\mathcal{A}$.

As was the case with mixing and ergodicity, reversing time has no effect. That is, if T is an isomorphism, $h(T) = h(T^{-1})$. In fact, we have

5.5 Theorem. (i) If $\mathcal{A}$ is a finite partition and $\bigvee_{-\infty}^{\infty} T^k \mathcal{A}$ generates Σ then $h(T) = h(\mathcal{A}, T)$. (See the definition of a K-system, 5.7).

(ii) If T^t is a stationary, measure preserving flow on X then

$$h(T^t) = |t|\, h(T^1).$$

Here $h(T^1)$ is called the entropy per second of the flow.

For the proofs and further properties, see Billingsly [1, p. 81-7] and Wightman [1].

The next aim is to examine briefly the notion of a K-system (Kolmogorov system). These are of basic importance and possess all the nice properties of the preceding sections. Before doing this we recall the proper way to ignore sets of measure zero.

5.6 Definition. Let (X, Σ, μ) be a measure space. For $A, B \subset X$, let $A+B = (A\backslash B) \cup (B\backslash A)$, called the symmetric difference. Let R denote the equivalence relation ARB iff $\mu(A+B) = 0$, and $\mathcal{B} = \Sigma/R$ the collection of equivalence classes, called the measure algebra.

By abuse of language we often write a representative for the class, and write 'mod 0' to indicate this.

The measure algebra inherits most of the structure of X. For example we can form $A \cup B$ (mod 0), extend μ to $\mathcal{B}$ etc. See Halmos [1, p. 42-45]. For a penetrating discussion of measure algebras see Halmos-von Neumann [1].

5.7 Definition. Let T^t be a stationary, measure preserving flow on a finite measure space (X, Σ, μ). Then T^t is called a K-system (K-flow) iff there is a σ-algebra $\mathcal{A} \subset \Sigma$ (sub σ-algebra) such that

(i) $\mathcal{A} \subset T^{-t}\mathcal{A}$ if $t \geq 0$ (mod 0).

(ii) $\bigcap_{-\infty}^{\infty} (T^{-t}\mathcal{A}) = \{\emptyset, X\}$ (mod 0).

(iii) $\bigvee_{-\infty}^{\infty} (T^{-t}\mathcal{A}) = \Sigma$ (mod 0).

Here $T^{-t}\mathcal{A} = \{T^{-t}A : A \in \mathcal{A}\}$ and $\bigvee_{-\infty}^{\infty} T^{-t}\mathcal{A}$ is the σ-algebra generated by all $T^{-t}\mathcal{A}$.

The definition of K-cascades is similar.

For an equivalent definition in terms of partitions see Sinai [1, p.85].

Note that from (i) we have $T^{-t}\mathcal{Q} \subset T^{-s}\mathcal{Q}$ if $t \le s$ since $T^{-t}\mathcal{Q} \subset T^{-s}\mathcal{Q}$ iff $\mathcal{Q} \subset T^{t-s}\mathcal{Q}$, so that $T^{-s}\mathcal{Q} = \bigvee\{T^{-t}\mathcal{Q} : t \le s\}$. Also, from this and (iii), $T^{-t}\mathcal{Q}$ increases to Σ as $t \longrightarrow \infty$ (equivalent to (iii)).

Roughly, (i) says that $\mathcal{Q}$ is being scrambled, (ii) that there was no information at $-\infty$ $(H\{\emptyset, X\} = 0)$ and (iii) that the information, or randomness is maximal at $t = +\infty$.

The key theorem of K-systems is:

5.8 <u>Theorem</u>. (Sinai-Kolmogorov). Let T^t be a K-system on the probability space (X, Σ, μ) with propagator U^t on $L^2(X,\mu)$. (2.4). Then we have

(i) $h(T^1) > 0$ (5.4).

(ii) T^t is ergodic and mixing of all orders (see 5.9 below).

(iii) the spectrum of U^t consists of the simple eigenvalue 1 (corresponding to constants), together with an infinite number of copies of Lebesgue measure on the circle (see 2.11).

For the proof, see Sinai [1, p. 104-8]. Actually, the proof that T^t is ergodic is trivial. In fact, if $T^{-t}A = A$ for all t, we have $A \in T^{-t}\mathcal{Q}$ for some t, and hence all t, (by (iii)). Hence, by (ii), $\mu(A) = 0$ or $\mu(A) = 1$, so T^t is ergodic. The mixing properties seem a little more subtle. (However, using Halmos [1, p. 39] we get weak mixing.)

5.9 <u>Definition</u>. Let T^t be a measure preserving flow on a probability space (X, Σ, μ). T^t is called <u>mixing of order</u> n iff for all $A_1, \ldots, A_n \in \Sigma$,

$$\lim_{\inf|t_i - t_j| \to \infty} \mu(T^{-t_1} A_1 \cap \dots \cap T^{-t_n} A_n) = \mu(A_1) \dots \mu(A_n) .$$

Mixing of order one is measure preserving and of order two is usual mixing. Taking $A_n = X$, we see mixing of order n implies mixing of order $n-1$.

As with ergodicity and mixing, the notion of a K-system is insensitive to the direction of time.

5.10 <u>Theorem</u>. Suppose T^t is a K-system and $S^t = T^{-t}$. Then S^t is also a K-system.

<u>Proof</u>. If $\mathcal{A}$ is the σ-algebra associated with T^t, let $\mathcal{B}$ be the σ-algebra generated by $\Sigma \setminus \mathcal{A}$. But $T^{-t}(\Sigma \setminus \mathcal{A}) \subset \Sigma \setminus \mathcal{A}$ so that $T^{-t}\mathcal{B} \subset \mathcal{B}$, or $S^{-t}\mathcal{B} \supset \mathcal{B}$. To check (ii) of 5.7, $\bigcap_{-\infty}^{\infty} T^t(\Sigma \setminus \mathcal{A}) = \Sigma \setminus \bigvee_{-\infty}^{\infty} T^t \mathcal{A} = \emptyset$ from which the result follows, (iii) being similar. $\square$

We turn next to the notion of a C-system (C-flow) (sometimes called a U-system). The history of C-systems is a long and involved one, beginning with special examples of Hadamard and Hopf on flows on (compact) two dimensional manifolds with negative curvature. Geodesic flows in these cases are ergodic, mixing and even K-systems. (See Anosov [1]). The key geometrical properties were isolated by Anosov.

To prepare the definition we recall a few more facts from differential geometry.

5.11 <u>Definition</u>. Let M be a manifold and $m \in M$. A <u>curve at</u> m is a smooth map $c: (-a,a) \longrightarrow M$ with $c(0) = m$, $a > 0$. Two curves

c_1, c_2 at m are called equivalent iff in some (and hence every) local chart, $Dc_1(0) = Dc_2(0)$ (are tangent at the origin). The equivalence classes of curves at m form the tangent space at m, denoted $T_m M$. The tangent space of M is defined by $TM = \cup\{T_m M: m \in M\}$, and the canonical projection $\tau: TM \longrightarrow M$ by $\tau(T_m M) = m$.

If M has dimension n, it is easy to see that $T_m M$ has the natural structure of an n-dimensional vector space and TM that of a $2n$-dimensional manifold.

If $f: M \longrightarrow N$ is a map of manifolds, its tangent is $Tf: TM \longrightarrow TN$, given by $Tf(c) = f \circ c$ (equivalence class at $f(c(0))$). In local coordinates Tf is just the derivative of f.

The chain rule of calculus becomes $T(f \circ g) = Tf \circ Tg$.

For proofs of these facts, see Abraham [1].

5.12 Definition. A vector field on a manifold M is a smooth map $X: M \longrightarrow TM$ such that $\tau \circ X$ is the identity on M (attaches a vector to each point). A curve $c: (-a,a) \longrightarrow M$ is called an integral curve of X iff $\frac{dc}{dt}(t) = X(c(t))$ where $\frac{dc}{dt}(t)$ is the usual derivative ($Tc(t) \cdot 1$ in terms of T). A flow T^t on M is called the integral of X iff each orbit of the flow is an integral curve. We sometimes speak of X as possessing the flow.

The basic existence and uniqueness theorem for ordinary differential equations yields:

5.13 Theorem. Let M be a compact manifold and X a (smooth) vector field on M. Then X possesses a unique smooth flow.

See Abraham [1, §7]. If M is not compact we obtain in general only a local flow.

5.14 Definition. A Riemannian manifold is a manifold M together with a smooth map $g\colon\ M \longrightarrow T^2_0(M)$ with $\tau^2_0 \circ g$ the identity, where $T^2_{0m}(M)$ consists of all real bilinear maps

$$b\colon\ T_m(M) \times T_m M \longrightarrow R \quad \text{and} \quad T^2_0(M) = \cup\{T^2_{0m}(M)\colon\ m \in M\}$$

(which can be given the structure of a manifold), and again $\tau^2_0\colon\ T^2_0(M) \longrightarrow M$ is the projection, and moreover, for each $m \in M$ we require g to obey:

(i) $g(m)(v,v) \geq 0$ and $= 0$ iff $v = 0$ for all $v \in T_m M$.

(ii) $g(m)(v,w) = (w,v)$ for all $v,\ w \in T_m M$.

If (i) is replaced by the weaker condition

(i)' $g(m)(v,w) = 0$ for all $w \in T_m M$ implies $v = 0$

then g is called a pseudo-Riemannian metric.

Roughly speaking, g is the smooth assignment of an inner product to each tangent space of M . Each tangent space then has a norm $\|v\|_m = g(m)(v,v)^{1/2}$.

We are now ready to define C-systems.

5.15 Definition. Let M be a compact Riemannian manifold and X a (smooth) vector field on M with flow T^t (by 5.13). Let $\tilde{T}^t = T(T^t)\colon\ TM \longrightarrow TM$ (which is a flow on TM by the chain rule) (in fact, $\tilde{T}^t$ is the flow of $TX\colon\ TM \longrightarrow T(TM)$, a "second order equation"; see Abraham [1, §17]). We say that X (or T^t) is a C-system (or C-flow) iff the following conditions hold:

(i) X has no critical points; that is, $X(m) \neq 0$ for all $m \in M$.

(ii) for each $m \in M$ we can write

$$T_m M = T_{1m} M \oplus T_{2m} M \oplus T_{3m} M$$

where

(a) $T_{3m}M$ is the subspace generated by $X(m) \in T_m M$, which by (i) is therefore one dimensional.

(b) $T_{1m}M$ is at least one dimensional, and for each $v \in T_{1m}M$, we have

$$\|\tilde{T}^t v\|_{T^t(m)} \leq a\|v\|_m \, e^{-ct} \quad \text{for} \quad t \geq 0$$

$$\|\tilde{T}^t v\|_{T^t(m)} \geq b\|v\|_m \, e^{-ct} \quad \text{for} \quad t \leq 0$$

where a, b, c are positive constants independent of v .

(c) $T_{2m}M$ is at least one dimensional, and for each $v \in T_{2m}M$ we have

$$\|\tilde{T}^t v\|_{T^t(m)} \leq a\|v\|_m \, e^{ct} \quad \text{for} \quad t \geq 0$$

$$\|\tilde{T}^t v\|_{T^t(m)} \geq b\|v\|_m \, e^{ct} \quad \text{for} \quad t \leq 0 .$$

It follows easily that the subspaces in (ii) are uniquely determined and depend continuously on m. See Avez [2, p. 5].

Roughly speaking, the situation is as follows: in one subspace T_2, we have orbits of tangent vectors expanding exponentially along the base orbit $T^t(m)$ while in the other, they are decaying.

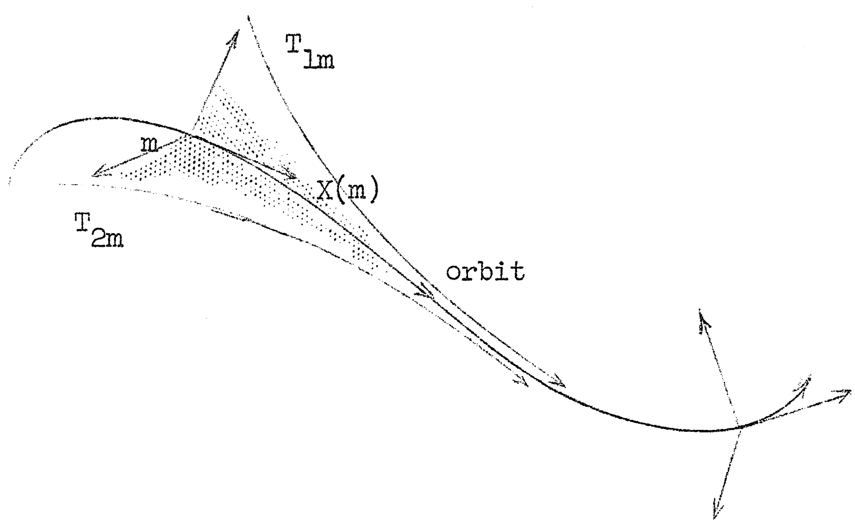
Notice that if a vector field possesses a closed orbit (an orbit for which $T^{\tau+t}(m) = T^t(m)$ for some $\tau > 0$) then it cannot be a C-system, for this would imply that $\|v\|_m = 0$ or ∞.

Also note that the direction T_3 is essential since vectors here are always constant multiples of X under the flow $(\tilde{T}^t X = X)$.

Sometimes $\tilde{T}^t v$ is written $T_*^t v$, (in Abraham [1]).

The notion of a C-system clearly makes no sense for cascades since we do not have any corresponding vector field in general. For the modification in case of a C-cascade, see Arnold-Avez [1, p. 47].

It is difficult to picture the situation globally since the manifold is compact and three dimensional (one requires at least R^4 in which to embed it), so we pretend we are looking locally at R^3:



The two basic theorems of Anosov on C-systems are as follows.

5.16 Theorem. Every C-system is either a K-system or has a non constant eigenfunction.

5.17 Theorem. C-systems are structurally stable.

The proof of the first theorem has yet to appear. However, see Avez [2, p. 85] for the proof that C-systems are ergodic. For the precise meaning of the second see Abraham [1]. Roughly it means that the property of being a C-system is retained under small perturbations of X, which is a great comfort to physicists for it means that small effects not considered will not crush the theory (which can happen!).

The proof of 5.17 (also due to Anosov) has recently been simplified by J. Moser and streamlined by J. Mather and R. Abraham (see appendix in Smale [1]). The proof is also outlined in Arnold-Avez [1, p. 55-60].

Unfortunately, C-systems are not generic, (a property is generic, roughly if almost every vector field possesses it), as shown by an example of Smale. (Appendix 24 of Arnold-Avez [1]).

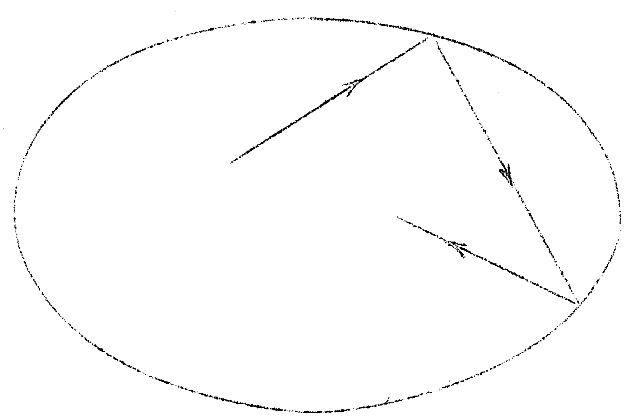
Sinai's procedure for dealing with hard spheres in a box is related to the ideas for a C-system, but of course can't use them since the flow is discontinuous (see §1). One possible procedure is to approximate the flow by C-flows and show that they are "uniformly mixing" or "uniformly K-systems" so that in the limit these properties are not destroyed.

There is a useful criterion for a flow to be a C-flow in terms of curvature:

5.18 Theorem (Hadamard-Cartan). Let M be a compact connected Riemannian manifold with negative curvature. Then the geodesic flow on the unitary tangent bundle of M is a C-flow, and a K-flow.

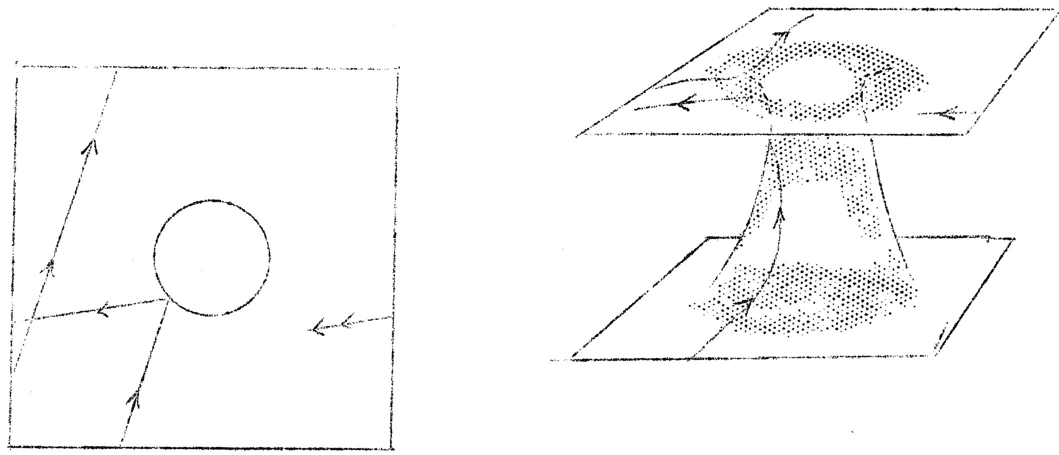
This is proven in Arnold-Avez [1, app. 21]. Let us explain the terms in 5.18. First, a topological space X is connected iff no subset other than $\emptyset$, X is both open and closed. In the case of manifolds this is equivalent to: any two points can be joined by a continuous curve. (See Abraham [1, app. A]). Compact was explained in 1.8. For the definition of curvature of a Riemannian manifold, see Helgason [1, Ch. I, §8, 9, 13]. The geodesic flow may be thought of as the motion of a free particle with kinetic energy given by g . This is explained in detail in Abraham [1, §18]. By conservation of energy the length of the tangent vectors is preserved under the flow so it makes sense to talk about the unitary tangent bundle; that is the submanifold consisting of tangent vectors of length one.

For further intuition we consider two examples (Arnold-Avez [1, p.65-67]). First, consider playing billiards on an elliptical table. The flow of a typical point is shown below.



This can be thought of as geodesic motion on a flattened ellipsoid (which has positive curvature). The flow is not ergodic, principally because of the 'focusing property' of the ellipse.

Next consider a torus with an obstruction.



Here the motion can be thought of as geodesic motion on a surface of negative curvature which has been flattened, although for visualization we require two sheets of a torus. The negative curvature is all concentrated around the obstruction. Here the flow is a limiting C-system and so is ergodic. The reason is the defocusing effect (scattering) of the obstruction.

For a rigorous description of these examples one uses generalized tensor analysis (distributions) and generalized Hamiltonian systems (generalized geodesic flows).

For further discussion along these lines, see Wightman [1] and Arnold-Avez [1].

§6. The Virial Theorem and Transport Properties.

In this section we prove some theorems in statistical mechanics to illustrate the application of the ergodic theorems.

We shall temporarily suspend the procedure of including all the background material, for here some proofs (6.2, 6.4, 6.5) require a good knowledge of mechanics. We shall summarize the notations, however. That assumed is Abraham [1, Ch. I-III]. The treatment is global and is valid in general relativity, for example.

It is useful to illustrate and interpret the results by a standard Hamiltonian in flat space R^{6N}, given by

$$H(x,p) = \sum_{j=1}^{N} \frac{p_j^2}{2m_j} + \sum_{j=1}^{N} V_j(x_j) + \sum_{j<k} V_{jk}(x_j - x_k)$$

where $x, p \in R^{3N}$, $x_j, p_j \in R^3$ being the 'components'. Here V_j is a potential acting on individual particles (one particle potential) and V_{jk} represents the interparticle forces (two particle potential). Of course such a Hamiltonian is special to a flat space. Readers unfamiliar with global mechanics should adapt special proofs for this case.

6.1 Summary of Notation.

(1) T^*M denotes the cotangent bundle of a manifold M, defined as follows. For each $m \in M$, T_m^*M is the collection of R-linear maps $\alpha: T_mM \longrightarrow R$ (see 5.11), that is, the dual space to T_mM. Then $T^*M = \cup\{T_m^*M:\ m \in M\}$. T^*M has the natural structure of a manifold (or more generally, a vector bundle).

(2) $\mathcal{F}(M)$ denotes the smooth real valued functions on M; $f: M \longrightarrow R$.

(3) d denotes the exterior derivative; it maps k forms into $k+1$ forms.

(4) θ denotes the canonical one form on T^*M. It is a mapping θ: $T^*M \longrightarrow T^*(T^*M)$ defined by $\theta(\alpha_m)\cdot v_{\alpha_m} = -\alpha_m(T\tau_M^*\cdot v_{\alpha_m})$ where $\alpha_m \in T_m^*M$, $v_{\alpha_m} \in T_{\alpha_m}(T_m^*M)$ and τ_M^*: $T^*M \longrightarrow M$ is the canonical projection. In terms of local coordinates, $\theta = \sum_{i=1}^{n} q^i dp_i$. See Abraham [1, §14]. $\omega = d\theta$ is the symplectic form on T^*M.

(5) P_X denotes the momentum of a vectorfield X on M. $P_X \in \mathcal{F}(T^*M)$ and is defined by $P_X(\alpha_m) = \alpha_m\cdot X(m)$. For example, if X is $\frac{\partial}{\partial q^i}$, P_X is p_i, (linear momentum) and similarly for angular momentum. P_X enters into classical mechanics in a fundamental way when we deal with symmetry groups and conservation laws.

(6) $\{f,g\}$ denotes the Poisson bracket of two smooth functions. In local coordinates it is the usual expression.

(7) X_H denotes the Hamiltonian vectorfield of a function $H \in \mathcal{F}(M)$ where M is a symplectic manifold (say T^*M above). It is obtained from dH by means of the symplectic form ω. The flow of X_H is the motion of the system. A basic fact is that if T^t is the flow,

$$\frac{d}{dt}(f\circ T^t) = \{f,H\}\circ T^t$$

(equations of motion). This is also denoted $L_{X_H} f$; L_X denoting the Lie derivative.

(8) If $\varphi\colon\ M \longrightarrow N$ is a diffeomorphism that is, is smooth, one to one and onto, with φ^{-1} smooth, φ^* denotes the corresponding map of tensors, and forms.

The first theorem we consider is generally referred to as equipartition of energy.

6.2 Theorem. Suppose M is a symplectic manifold, $H \in \mathcal{F}(M)$ and $\Sigma_e = H^{-1}(e)$ is a compact manifold. Equip Σ_e with an orientation (volume) Ω_e invariant under the flow of X_H induced on Σ_e, and suppose the flow is ergodic (to Ω_e corresponds a measure). Suppose further that there are functions $f_1, f_2 \in \mathcal{F}(\Sigma_e)$, $f_i(m) > 0$ for all $m \in \Sigma_e$ and a diffeomorphism $\varphi\colon\ \Sigma_e \longrightarrow \Sigma_e$ such that $\varphi^*(f_1\Omega_e) = f_2\Omega_e$. Then

$$\hat{f}_1 = \hat{f}_2 \quad \text{a.e.}$$

where $\hat{f}_1$ denotes the (constant) time average of f_1. (See 3.3). The same theorem holds if we merely assume f_1 and f_2 are integrable.

Proof. That Σ_e admits an invariant volume and we get a flow on Σ_e is proven in Abraham [1]. Since f_1 and f_2 are positive, φ is orientation preserving, and by the change of variables theorem (Abraham [1, §12]),

$$\int f_1 \, d\mu = \int f_2 \, d\mu$$

where μ is the measure on Σ_e. Hence we have the result by the Birkhoff ergodic theorem 3.3. (The change of variables formula is easily extended to L^1 functions.) $\square$

6.3 Corollary. On R^{6N} consider a Hamiltonian of the form

$$H(x,p) = \sum_{j=1}^{N} \frac{p_j^2}{2m_j} + V(x_1,\ldots,x_N)$$

and suppose $H^{-1}(e)$ is compact and the flow is ergodic on $H^{-1}(e)$. Then the time averages of $\frac{p_j^2}{2m_j}$ and $\frac{p_k^2}{2m_k}$ are equal, as well as their individual components.

Proof. In 6.2 take $f_1 = p_j^2/2m_j$, $f_2 = p_k^2/2m_j$ and let $\varphi\colon \Sigma_e \longrightarrow \Sigma_e$ be given by interchanging $p_j/\sqrt{m_j}$ and $p_k/\sqrt{m_k}$. By symmetry, the conditions of 6.2 apply. $\square$

Similar corollaries can be derived if V has a special form; say quadratic in x_j. See Huang [1, p. 149].

It is via 6.3 that the temperature of a system is defined. Each 'degree of freedom' is assigned a value $\frac{1}{2}kT$ where k is a constant. Thus the temperature of the system is defined so the time average of $\sum_{j=1}^{N} p_j^2/2m_j$ is $\frac{3}{2}NkT$. Thus for each energy surface, T is some constant. Ergodicity is essential to make this a meaningful statement.

The next theorem is a generalization of the classical virial theorem. We deal with the case of a Hamiltonian derived from a pseudo-Riemannian metric (the Lorentz metric, or Euclidean metric for example). Incidentally, in case g is a Lorentz metric at each point, the parameter t obtained from the flow T^t is called the proper time. If the energy H is invariant under the action of the Lorentz group, so is the flow (Lorentz invariance). See Abraham [1, §§18, 22].

6.4 <u>Theorem</u>. Let M be a manifold, T^*M its cotangent bundle with the natural symplectic structure. Let $H = T + V$ where $V \in \mathcal{F}(M)$ and $T(\alpha_m) = \frac{1}{2}g(m)\cdot(\alpha_m,\alpha_m)$ where g is a pseudo-Riemannian metric. Suppose $H^{-1}(e) = \Sigma_e$ is a compact submanifold of T^*M. For any vectorfield X on M define $G_X \in \mathcal{F}(T^*M)$ by

$$G_X(\alpha_m) = dT(\alpha_m)\cdot X_{P_X}(\alpha_m) + dV(m)\cdot X(m) .$$

Then

(i) the time average of G_X on Σ_e is zero.

(ii) the space average of G_X on Σ_e is zero.

(iii) if the flow (of X_H) is ergodic on Σ_e the time and space averages coincide (termwise).

Moreover, in local coordinates (x,p) the function G_X is given by

$$G_X(x,p) = - g(x)\cdot(p,p\cdot DX(x)) + \frac{1}{2} D_x g(x)\cdot(p,p)\cdot X(x) + DV(x)\cdot X(x)$$

where D denotes derivative. As above, P_X is the momentum of X and X_{P_X} is the vectorfield associated to P_X.

For the proof, we first establish the following

6.5 <u>Lemma</u>. Let M be a symplectic manifold, $H \in \mathcal{F}(M)$ and $H^{-1}(e)$ be compact and $i\colon H^{-1}(e) \longrightarrow M$ the inclusion. For any $f \in \mathcal{F}(M)$, we have

(i) the time average of $\{f,H\}\circ i$ on $H^{-1}(e)$ is zero;

(ii) $\int \{f,H\}\circ i \, d\mu = 0$.

(Compactness is essential as simple examples show).

Proof. By abuse of notation we work directly on Σ_e. For (i), $\{f,H\}\circ T^t = \frac{d}{dt}(f\circ T^t)$ so that $\frac{1}{t}\int_0^t \{f,H\}\circ T^s\, ds = \frac{1}{t}(f\circ T^t - f) \longrightarrow 0$ as $t \longrightarrow \infty$ since f is bounded ($H^{-1}(e)$ is compact). For (ii), $\{f,H\} = L_{X_H} f$ and since Ω_e is invariant under the flow;

$\{f,H\}\,\Omega_e = L_{X_H}(f\Omega_e) = di_{X_H}(f\Omega_e)$ since $L_Y = di_Y + i_Y d$ and $d(f\Omega_e) = 0$.

But since Σ_e is compact, we have by Stokes theorem, $\int d(i_{X_H} f\Omega_e) = 0$, proving the assertion. Note that ergodicity is not required. $\square$

For example, if on R^{6N}, H is given as in 6.3, the time and space average of each p_j is zero since $p_j = \{q_j,H\}$. Thus, if $p_j > 0$, for example, the energy surface cannot be compact.

Proof of 6.4. We claim that $G_X = \{H,P_X\}$ so the result will follow from the lemma. To see this, we have

$$\{H,P_X\} = L_{X_{P_X}} T + L_{X_{P_X}} V$$

$$= dT\cdot X_{P_X} + dV\cdot X_{P_X} .$$

But the flow of X_{P_X} is induced by T^{t*} which projects to T^t on M; hence since V depends only on M; $dV\cdot X_{P_X} = dV\cdot X$. See also Sternberg [1, p. 146-7]. In local coordinates

$$X_{P_X}(x,p) = (X(x), -p\cdot DX(x))$$

obtained by differentiating T^{t*}. The second formula for G_X now follows, using bilinearity and symmetry of g . (iii) follows from the Birkhoff ergodic theorem 3.4. □

The Virial theorem 6.4 differs from the usual formulation by the extra term $D_x g$, which is interpreted as an effective potential due to the curvature of the space.

The following corollary is of special interest in statistical mechanics.

6.6 Corollary. On R^{6N} consider the Hamiltonian

$$H(x,p) = \sum_{j=1}^{N} \frac{p_j^2}{2m_j} + \sum_{j=1}^{N} V_j(x_j) + \sum_{j<k} V_{jk}(x_j - x_k)$$

where V_j and V_{jk} $j,k = 1,\ldots,N$ are smooth functions on R^3 . Suppose $H^{-1}(e)$ is compact so that we have a flow induced on $H^{-1}(e)$. Then

(i) the time average of the following quantity is zero:

$$-2 \sum p_j^2/2m_j + \sum_{j=1}^{N} \nabla V_j(x_j)\cdot x + \sum_{j<k} \nabla V_{jk}(x_j - x_k)\cdot(x_j - x_k)$$

(ii) the space average (over $H^{-1}(e)$) of the function in (i) is zero.

(iii) if the flow is ergodic on $H^{-1}(e)$, we have (equation of state)

$$p|V| = NkT - \frac{1}{3} \sum_{j<k} \int (x_j - x_k)\cdot\nabla V_{jk}(x_j - x_k)\,d\mu$$

(integration over $H^{-1}(e)$) where kT was defined above (6.3) and where

$$3p|V| = \sum_{j=1}^{N} \int \nabla V_j(x_j)\cdot x_j = \lim_{t\to\infty} \frac{1}{t}\int_0^t \nabla V_j(x_j(s))\cdot x_j(s)ds$$

by definition.

Proof. This is a special case of 6.4 using $X(x) = x$. □

Here $|V|$ is the volume of the 'box'

$$B = \{x \in R^3:\ x \text{ is a space coordinate of a point in } H^{-1}(e)\} .$$

If other forces are added to the box forces V_j , [such as 'gravity'; $m_j g\cdot x$] other terms must be added [$g\cdot\Sigma\int m_j x_j d\mu$ (center of mass)].

The constant p is called the pressure and corresponds to time average momentum transfer as indicated. In 6.6 almost every surface $H^{-1}(e)$ is a submanifold (Sard's theorem).

The theorem 6.5, 6.6 can be generalized to the case in which V_j , V_{jk} are not smooth (say are distributions) but the system has a flow. This requires distribution theory which we shant go into here. (The proof will appear elsewhere).

Finally in this section we briefly discuss transport properties to illustrate some of the difficulties present and what modifications of previous ideas may be necessary. This subject is in its infancy and represents a real challenge for any prospective workers in the field.

The point is to obtain macrocosmic information from microscopic information; for example the model of a gas such as hard spheres. The properties we have in mind are heat conduction and viscosity.

One of the problems is that in such models, energy is not conserved, but is rather transferred through the system. In that case we make use of a more general ergodic theorem as follows:

6.7 Theorem (Chacon-Ornstein [1]). Let T be a positive linear operator on L_1 of a measure space (X,Σ,μ) (that is, $f \geq 0$ implies $Tf \geq 0$) and suppose $\|Tf\|_1 \leq \|f\|_1$ for all $f \in L_1(X,\mu)$ (see 2.1). Then for each $f,g \in L_1(X,\mu)$, $g \geq 0$,

$$\lim_{n\to\infty} \left(\sum_{k=0}^{n} T^k f\right)/\left(\sum_{k=0}^{n} T^k g\right)$$

exists and is finite almost everywhere on the set

$$A = \{x \in X: \ T^k(g)(x) > 0 \ \text{ for some } \ k \geq 0\} .$$

In particular if $\mu(X) < \infty$, and T is induced by a point transformation,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} T^k f$$

exists almost everywhere on X for each $f \in L_1(X,\mu)$, (take $g = 1$).

There are other ergodic theorems which may be useful in a future development of the subject. We refer especially to Ackolou [1] and Ackolou-Sharpe [2].

To see why this theorem is appropriate, consider a model for heat conduction:

6.8 Definition. (Heat conduction model for a hard sphere gas in two dimensions). Let S^1 be the unit $I = [0,1]$ and $C = I \times S^1$ a cylinder. Let $C^n = C \times \dots \times C$ and $P = C^n \times R^{2n}$ with the set $\{(q,p): \ |q_i - q_j| \leq a \ \text{ for some } \ i,j\}$ removed (corresponding to collisions). Consider the flow on the phase space P described as follows

(for the flow to be well defined further sets corresponding to triple collisions, etc., must be removed)

(i) free motion between collisions

(ii) elastic collisions between the n spheres

(iii) "reflection" at the top wall $\{1\} \times S^1$ described by $(p_x, p_y) \longmapsto (P_x, \alpha p_y + (1-\alpha)p)$ for $p > 0$ and $0 < \alpha < 1$ constants;

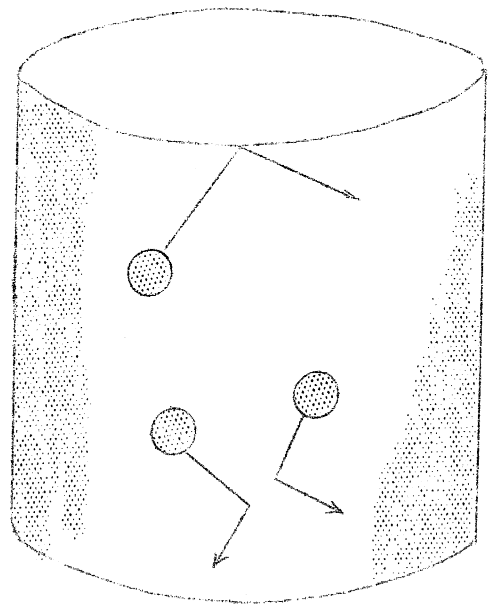
(iv) "reflection" at the bottom wall $\{0\} \times S^1$ described by $(p_x, p_y) \longmapsto (p_x, \beta p_y + (1-\beta)h)$ for $h > 0$ and $0 < \beta < 1$ constants.

6.9 Theorem. The above flow is a contraction. That is, it decreases the measure on P. Alternatively, the induced flow on L_1 has norm ≤ 1. $(\|Tf\| \leq \|f\|)$.

Proof. The measure on P is conserved except possibly under a reflection. It is sufficient to consider a single particle undergoing reflection at, say, the top wall. For a fixed time, t_o sufficiently small, the flow is of the form

$$(x, y, p_x, p_y) \longmapsto (x + t_o p_x,\ y + f(x, p_x, p_y),\ p_x,\ \alpha p_y + (1-\alpha)p)$$

which has Jacobian α; since $0 < \alpha \leq 1$, volume is decreased. $\square$



6.10 Conjecture. Almost every orbit above lies in some compact set (is bounded).

From 6.9 and 6.8 we can state the following (6.10 is not needed):

6.11 Definition (Thermal conductivity). In the above flow, the thermal conductivity κ is defined as follows: Let φ be a smooth function on C and let

$$-\kappa(\varphi) = \lim_{t\to\infty} \frac{\int_0^t \left[\sum_{i=1}^n \varphi(T^{-t}q_i)(T^{-t}p_i^2/2m)T^{-t}p_{iy}\right]dt}{\int_0^t \left[\sum_{i=1}^n \left(-\frac{\partial\varphi}{\partial y}\right)(T^{-t}q_i)\cdot T^{-t}(p_i^2/2m)\right]dt}$$

($\kappa(\varphi)$ depends on the initial point (q,p).)

Here the numerator represents the heat flux J_y in the y-direction, and the denominator represents the temperature gradient $\frac{\partial T_y}{\partial y}(\varphi) = T_y\left(-\frac{\partial\varphi}{\partial y}\right)$. Just from 6.7, we don't know the above limit exists as the denominator is not positive. We could however take the ratio of each with the total temperature, for example. The situation is less burdensome if the flow is bounded so that the time averages exist separately (see 6.7).

Some open problems for the above flow which would be worth investigation are as follows:

1. In what sense is the flow ergodic? (In the language of Ackolou-Sharpe [2], describe the boundary).

2. Is κ constant a.e.?

3. How does κ vary with the molecular radius a ?

4. Does T obey the heat equation in any reasonable sense (e.g.: in the thermodynamic limit §§7, 9).

(It may be that the above model is too complicated to warrant satisfactory analysis, and that the proper setting is the continuum limit.)

Viscosity seems to be a more difficult task than heat conduction, and there is, as yet no simple satisfactory model.

Until recently, the standard discussion of transport coefficients was based on Bogoliubov's generalization of the Chapman-Enskog theory, which leads to a power series expansion of the transport coefficients in the density. (For a review up until 1960, see Uhlenbeck and Ford [1] ch VII.) However, it was recently discovered that the series diverges, so that the entire subject needs rethinking. (For a recent account and references, see Dorfman and Cohen [1].)

For viscosity, the experimental situation the model should mimic is as follows: One observes the torque on one of a pair of concentric cylinders when the other is rotated at constant angular velocity, with the gas between. The temperature is stabilized by means of cooling apparatus for the two shells.

If the flow is laminar and governed by the Navier-Stokes equations, the viscosity is given by

$$\eta = \text{average x-momentum transferred in y direction} \div \text{y-gradient of x-momentum}$$

$$= \frac{L}{4\pi\Omega}\left(\frac{1}{a_1^2} - \frac{1}{a_2^2}\right)$$

where Ω is the angular velocity, a_1, a_2 the radii of the cylinders, and L the torque per unit length. (See Page [1, p. 278-80]).

A possible model would be similar to that for 6.8 with an x-component change as well at the top. Again this flow is a contraction. It is clear what the gradient of p_x in the y-direction should be, but the numerator is not so clear. Again we have many open problems. For example does the above formula for η hold in the thermodynamic limit? etc.

This field is fair game for research workers, but is probably a very difficult problem, on the level of Sinai's theorem for hard spheres in a box.

Helgason. Huang Frisch-Lebowitz.
Smale. Cohen.
~~Birkhoff~~
Sternberg.

BIBLIOGRAPHY

Abraham, R. and Marsden, J. [1] Foundations of Mechanics. Benjamin, New York (1967).

Anosov, D. V. [1] Ergodic Properties of Geodesic Flows on Closed Riemannian Manifolds of Negative Curvature. Dokl. Akad. Nauk SSSR 151 (1963) 1250-1252.

Apostol, T. [1] Mathematical Analysis. Addison Wesely Reading Mass. (1957).

Arnold, V. and Avez, A. Problemés Ergodique de la Mechanique Classique. Gauthier Villars Paris (1967).

Avez, A. Ergodic Theory of Dynamical Systems I, II Mimeographed. University of Minnesota (1966).

Berberian, S. [1] Measure and Integration. MacMillan, New York (1965).

Billingsly, P. [1] Ergodic Theory and Information. Wiley, New York (1965).

Birkhoff, G. D. [1] Proof of the Ergodic Theorem. Proc. Nat. Acad. 17 (1931) 656-660.

Doob, J. L. [1] Stochastic Processes. Wiley, New York (1953).

Dunford, N. and Schwartz, J. [1-2] Linear Operatiors I, II. Interscience, New York (1958).

Ehrenfest, P. and T. [1] The Conceptual Foundations of Statistical Mechanics. Cornell Univ. Press (1959).

Gibbs, J. W. [1] Elementary Principles of Statistical Mechanics. Collected Works, II Longmans Green and Co., New York (1928).

Halmos, P. R. [1] Lectures on Ergodic Theory. Chelsea, New York (1956).

Halmos, P. R. [2] Measure Theory. D. Van Nostrand, New York (1950).

Halmos, P. R. [3] Introduction to Hilbert Space and Spectral Multiplicity. Chelsea, New York (1951).

Halmos, P. R. and Von Neumann, J. [1] Operator Methods in Classical Mechanics (II). Ann. of Math. 43 (1942) 332-350.

Hopf, E. [1] Causality Statistics and Probability. J. Math. Phys. 13 (1934) (51-102).

Kolmogorov, A. N. [1] A New Metric Invariant of Transient Dynamical Systems and Automorphisms in Lebesque Spaces. Dokl. Akad. Nauk SSSR 119 (1958) 861-865.

Kolmogorov, A. N. [2] Entropy per Unit Time as a Metric Invariant of Automorphisms. ibid. 124 (1959) 754-755.

Kolmogorov, A. N. [3] General Theory of Dynamical Systems reprinted in appendix D of Abraham [1].

Koopman, B. O. [1] Hamiltonian Systems and Transformations in Hilbert Space. Proc. Nat. Acad. 17 (1931) 315-318.

Kouchnirenko, A. G. [1] An Estimate from above for the Entropy of a Classical Dynamical System. Sov. Math. Dokl. 6 (1965) 360-2.

Krylov [1] Works on the Foundations of Statistical Mechanics. Moscow (1950).

von Neumann, J. [1] Proof of the Quasi-Ergodic Hypothesis. Proc. Nat. Acad. 18 (1932) 70-82.

von Neumann, J. [2] Zur Operatorenmethode in der klassischen Mechanik. Ann. of Math. 33 (1932) 587-642, 789-791.

Shannon, C. E. [1] The Mathematical Theory of Communication. ed. W. Weaver, Univ. Ill. (1957).

Sinai, Ya. G. [1] Dynamical Systems with Countably-multiple Lebesque Spectrum. Izv. Akad. Nauk. SSSR 25 (1961) 899-924, A.M.S. transl. 39 (1964) (83-110).

Sinai, Ya. G. [2] Proof of the Ergodic Hypothesis for a System of Statistical Mechanics. Dokl. Akad. Nauk SSSR 153 (1963) 1261-1264.

Sinai, Ya. G. [3] On the Concept of Entropy for a Dynamical System. Dokl. Akad. Nauk. SSSR 124 (1959) 768-771.

Sinai, Ya. G. [4] Dynamical Systems with Countably Multiple Lebesque Spectrum II. Izv. Akad. Nauk 30 (1966) 15-68.

Stone, M. H. [1] Linear Transformations in Hilbert Space. Proc. Nat. Acad. 15 (1929) 198-200, 423-425; 16 (1930) 172-175.

Stone, M. H. [2] On One Paramater Unitary Groups in Hilbert Space. Ann. of Math. 33 (1932) 643-648.

Wightman, A. [1] Lectures on Statistical Mechanics. Notes (by G. Svethchny and A. Z. Capri) Mimeographed; Princeton (1966).
SVETLICHNY

Yosida, K. [1] Functional Analysis. Springer, Berlin (1965).

FURTHER REFERENCES FOR CHAPTER I

M. Ackolou [1] Pointwise Ergodic Theorems (to appear in trans. A.M.S.)

M. Ackolou and R. Sharp [2] Ergodic Theory and Boundaries (to appear)

G. D. Birkhoff [2] Dynamical Systems New York (1929)

E. Cartan [1] Lecons sur la géometrie des espaces de Riemann, Gauthier Villars, Paris (1928)

R. Chacon and D. Ornstein [1] A General Ergodic Theorem Ill. Journal Mathematics 4 (1960) 153

J. R. Dorfman and E. G. Cohen [1] Difficulties in the Kinetic Theory of Dense Gases. Journal Math. Phys. 8 (1967) 282-297.

I. M. Gelfand and S. V. Fomin [1] Geodesic Flows on Manifolds of Constant Negative Curvature Am. Math. Soc. Transl. 2 no 1, (1955) 49-67.

J. Hadamard [1] Les surfaces à courbures opposées et leur lignes géodèsiques. Journal Math. Pure et Appl. (5) 4 (1898) 27-73.

S. Helgason [1] Differential Geometry and Symmetric Spaces. Academic Press N. Y. (1962)

K. Huang [1] Statistical Mechanics, Wiley N.Y. (1962)

J. Marsden [1] Generalized Hamiltonian Mechanics (to appear).

L. Page [1] Introduction to Theoretical Physics Van Nostrand, Princeton, N. J. (1952)

Ya. G. Sinai [5] The central limit theorem for geodesic flows on manifolds of constant negative curvature. Sov. Math. Dokl. (1) 4 (1961) 983

[6] Geodesic flows on compact surfaces of negative curvature. Sov. Math. Dokl. (2) 1 (1961) 106

S. Smale [1] Differentiable dynamical systems (to appear in bulletin of A.M.S.)

S. Sternberg [1] Lectures on Differential Geometry Prentice Hall, Englewood Cliffs, N. J. (1964).

G. E. Uhlenbeck and G. W. Ford [1] Lectures in Statistical Mechanics. Am. Math. Soc. Providence R. I. (1963).

CHAPTER II THE THERMODYNAMIC LIMIT.

Roughly speaking, the thermodynamic limit is the passage from systems with a finite number of degrees of freedom to those with an infinite number, by letting the volume and number of particles tend to infinity, their ratio approaching a finite limit.

There are considerable mathematical difficulties associated with this limit (The C^* algebra approach avoids this by working directly in the limit, hopefully simpler). We begin in §7 with a rigorous treatment of the classical theorems of Yang and Lee. These are generalized in §9, along with associated problems (the limit in the other ensembles for example). In §10 we discuss the elegant treatment of correlation functions of Ruelle (with a few modifications and generalizations).

§7. The Classical Theory of Phase Transitions

This section contains a breif outline of the theories of Mayer and of Lees and Yang on phase transitions. This is mainly for orientation and motivation for later work. The theorems of Lee and Yang are proven in detail, as the proofs in currant literature are sketchy and omit a number of points, obscuring the real subtlety of the theorems. For the connection of the definitions with elementary physics we refer to Huang [1]. Many of the key papers in the subject are conveniently reprinted in Frisch-Lebowitz [1].

7.1 Definition. Let M be a phase space (symplectic manifold) and $H: M \longrightarrow R$ a Hamiltonian; not necessarily smooth. Let $D \subset M$ be a measureable set (recall that a symplectic manifold is canonically endowed with a measure). The partition function of H (relative to D) is defined by

$$Q_D: R^+ \longrightarrow R$$

$$Q_D(\beta) = \int_D \exp(-\beta H)\, d\mu \leq \infty$$

where $\beta \in R$, $\beta > 0$, and μ is the phase volume.

For each $\beta \in R$, $\beta > 0$, the function $e^{-\beta H}$ thought of as an (unnormalized) probability density function on M is called the canonical ensemble, so that $Q_D(\beta)$ is the probability for the event D.

From the physical point of view we think of $\beta = {}^1/kT$, T the temperature and k a constant connecting β to physical units (Boltzman's constant). The canonical ensemble then measures the distribution of states at a given temperature and is essentially justified by the central limit theorem.

We whall be interested in the following situation.

7.2 Theorem. Suppose $M = U \times R^{3n}$ where $U \subset R^{3n}$ is open, and $H(q,p) = \sum_{j=1}^{n} p_j^2/2m + V(q)$ for $q \in U$, $p \in R^{3n}$; $p = (p_1, \ldots p_n)$; $p_j \in R^3$ and $p_j^2 = <p_j, p_j>$ the usual inner product.

If $D = U_o \times R^{3n}$; $U_o \subset U$ is measureable, then

$$Q_D(\beta) = \frac{1}{\lambda^{3n}} \int_{U_o} \exp(-\beta V) dq$$

where $\lambda = (2\,\pi\beta/m)^{1/2}$. We often write Q_{U_o} for Q_D in this case

Proof. Here $\mu = \mu_q \times \mu_p$ where μ_q and μ_p are Lebesgue measure and $\times$ denotes the product measure (1.6). One of the basic properties of product measure is Fubini's theorem (Halmos [2, p.148]) which applies here as the integrands are positive. We obtain

$$Q_D(\beta) = \int_{U_o \times R^{3n}} \exp(-\beta\, \Sigma\, p_j^2/2m)\ \exp(-\beta V)(dq \times dp)$$

$$= \int_{R^{3n}} \exp(-\beta\, \Sigma\, p_j^2/2m)dp \int_{U_o} \exp(-\beta V)dq$$

The first integral is easily seen (i.e. is well known) to be $1/\lambda^{3n}$. □

The idea behind 7.1 is to allow for fluctuations in energy arising from weak interaction with a much larger system (temperature bath). The next concept allows fluctuations in the number of particles, arising from passage of particles into and out of the temperature bath."

7.3 Definition. Let M^o, M^1, M^2,... be symplectic manifolds and $H^n: M^n \longrightarrow R$ Hamiltonians. Let $D^n \subset M^n$ and define

$$\mathcal{Q}_D(z,\beta) = \sum_{n=o}^{\infty} \frac{z^n}{n!} Q_{D^n}(\beta)$$

for those pairs $\beta \in R$, $Z \in C$ for which the series converges. $\mathcal{Q}_D$ is called the grand partition function and for each z,β with $z \in R$, > 0, the function $\frac{z^n}{n!} Q_{D^n}(\beta)$ of n, thought of as a probability distribution is called the grand canonical ensemble.

The mean of a mapping $f: \{0,1,2,...\} \longrightarrow R$ is, as usual, defined by $(\operatorname{re} z, \beta, D^n)$

$$\langle f\rangle = \sum_{n=0}^{\infty} f(n) \frac{z^n}{n!} Q_{D^n}(\beta) \Big/ \sum_{n=0}^{\infty} \frac{z^n}{n!} Q_{D^n}(\beta)$$

In particular if n denotes the identity map, $\langle n\rangle$ is the "mean number of particles", and $\langle n\rangle = z \frac{\partial}{\partial z} \log \mathcal{Z}_D(z, \beta)$ (see below)

If $D^n = U^n \times R^{3n}$ we often write $\mathcal{Z}_{U^n}(z, \beta)$ for $\mathcal{Z}_{D^n}(z, \beta)$.

The expression for $\langle n\rangle$ is valid in any simply connected region where $\mathcal{Z}_D(z, \beta)$ does not vanish and is holomorphic (analytic), for then $\frac{\partial}{\partial z} \log \mathcal{Z}_D(z, \beta) = \frac{\partial}{\partial z} \mathcal{Z}_D(z, \beta) / \mathcal{Z}_D(z, \beta)$ and we may differentiate term by term. See Ahlfors [1, p. 139].

More general conditions for the convergence of $\mathcal{Z}_D$ are given in the next section.

Just as β was related to the temperature, z is related to the <u>chemical potential</u>.

We consider now, more specifically, two body forces;

7.4 <u>Definition</u>. Let $M^n = R^{3n} \times R^{3n}$ and suppose $H_n\colon M^n \longrightarrow R$ is given by

$$H_n(q,p) = \sum_{j=1}^{n} p_j^2/2m + \sum_{j<k} V_{jk}(|q_j - q_k|) + \sum_{i=1}^{n} V_i(q_i)$$

where $V_{jk}\colon R \longrightarrow R$ and $|\cdot|$ denotes euclidean length.

Suppose $U \subset R^3$ is open and $V_i = 0$ on U. Then we write $Q_U(\beta)$ or $Q_U^n(\beta)$ for $Q_{U\times\ldots\times U}(\beta)$ and $\mathcal{Z}_U(z,\beta)$ for $\sum_{n=0}^{\infty} \frac{z^n}{n!} Q_U^n(\beta)$. (Strictly speaking $U \times\ldots\times U$ may not be the configuration space as V_{ij} can be singular. However, we put $e^{-\beta V_{ij}} = 0$ if $V_{ij} = \infty$; also, V_i may represent a potential confining the particles to U; a δ function at the boundary for example.)

Thus $Q_U(\beta) = \frac{1}{\lambda^{3n}} \int_{U\times\ldots\times U} \exp(-\beta \sum_{i<j} V_{ij})dq.$

The <u>pressure</u> $P_U(z,\beta)$ is defined by

$$|U|\beta\, P_U(z,\beta) = \log \mathcal{Z}_U(z,\beta)$$

for suitable z (see above) where $|U|$ denotes the lebesgue measure of U in R^3 (volume).

The <u>specific volume</u> v is defined by

$$\frac{1}{v_U(z,\beta)} = \langle n\rangle/|U| = \frac{1}{|U|}\, z \frac{\partial}{\partial z} \log \mathcal{Z}_U(z,\beta)$$

again if it makes sense.

These two formulas are called the <u>parametric equations of state</u> in the grand canonical ensemble (compare §6).

The Ursell-Mayer theorem expresses Q and $\mathcal{Z}$ in terms of 'clusters'. For this, we prepare the following definition.

7.5 <u>Definition</u>. Consider the situation described in 7.4 and define

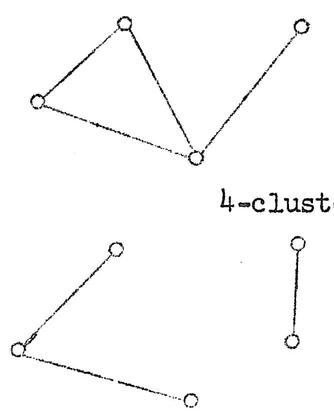
$f_{ij}: R \times R^3 \longrightarrow R$ by $f_{ij}(\beta, q) = \exp\{-\beta V_{ij}(|q|)\} - 1$ so that

$$Q_U(\beta) = \frac{1}{\lambda^{3n}} \int_{U\times\ldots\times U} \{\prod_{i<k} [1 + f_{jk}(\beta, q_i - q_k)]\} dq$$

Let $\ell \geq 1$ be an integer and $N_\ell = \{(j,k): j < k;\ 0 \leq j \leq \ell;\ 0 \leq k \leq \ell;\ j,k \text{ integers}\}$. An __$\ell$-cluster__ I is a subet of N_ℓ such that we cannot write $I = I' \cup I''$ where $I' \subset N_\ell$, $I'' \subset N_\ell$ and contain no common integers, (That is, I is a connected graph). The __ℓ-cluster coefficient__ is defined by

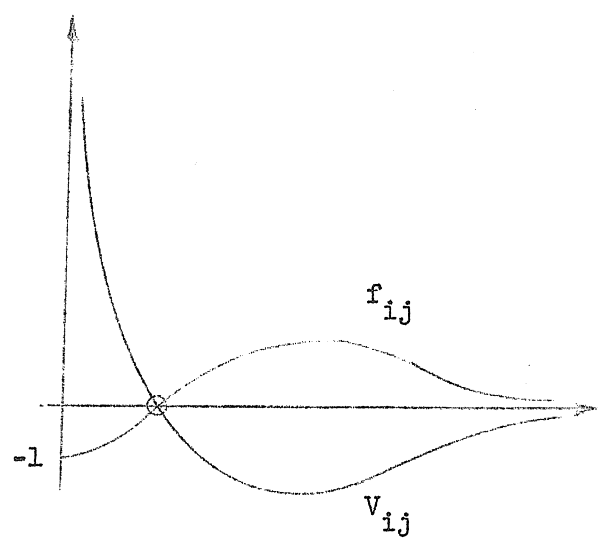
$$b_{\ell u}(\beta) = \frac{1}{\ell!\lambda^{3\ell-3}\,|U|} \Sigma\{\int f_I d\mu \ \ I \text{ is on } \ell\text{-cluster};\ f_I = \pi f_{ij};(i,j) \in I\}$$

where the integration is over as many copies of U as there are variables.



4-cluster

3-cluster & 2-cluster



7.6 <u>Theorem (Ursell-Mayer)</u> In 7.4, 7.5 we have

$$\frac{1}{n\,!}\, Q_U(\beta) = \Sigma \left\{ \prod_{\ell=1}^{n} \frac{1}{m_\ell !} \left[\frac{|U|}{\lambda^3}\, b_{\ell,U}(\beta)\right]^{m_\ell} ; \right.$$

$$\left. \sum_{\ell=1}^{n} \ell m_\ell = n; \quad m_1, \ldots, m_n \text{ positive integers} \right\}$$

and $$\frac{1}{|U|} \log \mathcal{Z}_U(z,\beta) = \frac{1}{\lambda^3} \sum_{\ell=1}^{\infty} b_{\ell,U}(\beta)\, z^{\ell}$$

The proof is a good exercise in combinatorics which we shall omit. See Huang [1, Ch. 14], for example.

This theorem is historically important as it was a method for handling the behavior of P and v as $|U| \longrightarrow \infty$, the thermodynamic limit (see below). They played a role in the Van der Waals equation for example, using the so-called virial expansion, which with cavalier mathematics is as follows:

$$\beta\, P_v = \Sigma\, A_\ell \left(\frac{\lambda^3}{U}\right)^{\ell-1}$$

where $b_{\ell,U} \longrightarrow b_\ell$ as $|U| \longrightarrow \infty$ and

$$\Sigma\, A_\ell (\Sigma\, n\, b_n\, z^n)^{\ell-1} = \Sigma\, b_\ell\, z^\ell / \Sigma\, \ell\, b_\ell\, z^\ell .$$

Mayer's idea in this program was that for small real $z > 0$ we should have a gas phase with all functions analytic. The first real $z_0 > 0$

for which $\frac{1}{v}$ fails to converge should correspond to a phase transition (condensation point). There are two main difficulties with this. First, we could run into the radius of convergence before a real singularity (there is no reason to believe that b_n are all real and positive). Secondly, any subsequent phase transitions are masked.

The above situation was markedly improved by two fundamental theorems of Lee and Yang (7.11 and 7.12 below). They studied a particular potential, described as follows:

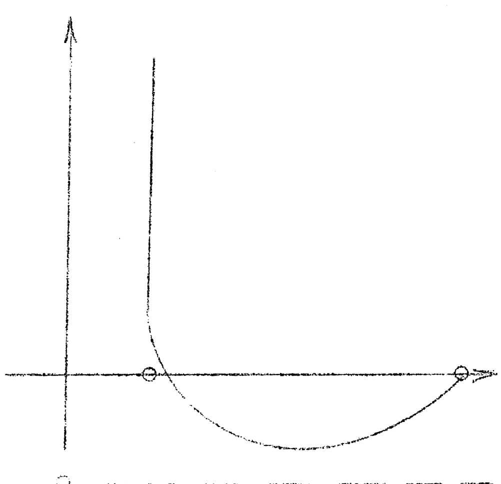
7.7 <u>Definition</u>. Consider the situation of 7.4. We say that $V_{ij} = V$ is a <u>Lee-Yang potential</u> iff the following conditions hold:

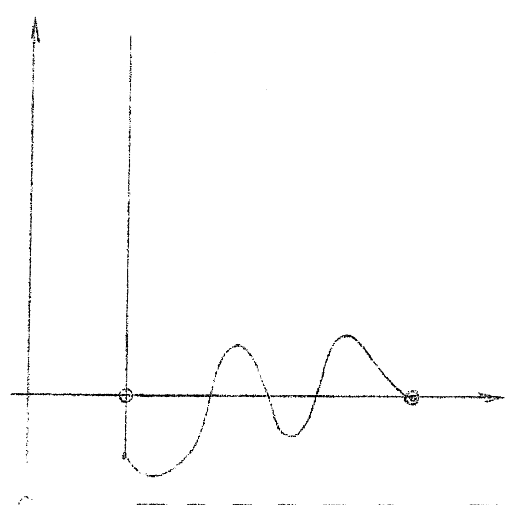
(i) $V(r) = \infty$ if $r \leq a$ where $a > 0$

(ii) there is a constant $B > 0$ such that $V(r) \geq -B$ for all $V \in R$

and (iii) $V(r) = 0$ for $r \geq r_0$ where $r_0 \geq a$.

Typical potentials are the following:





The crucial facts are that there be a non vanishing core ($a > 0$) a lower bound on the potential and finite range on the forces ($r_0 < \infty$). We see no reason to demand that $V(r) \leq 0$ for $a < r < r_0$ or that V be continuous.

The first important fact about the Lee-Yang potential is the following:

7.8 Lemma In 7.4, suppose $V_{ij} = V$ is a Lee-Yang potential and $|U| < \infty$. Then the grand partition function is a polynomial in z of degree $N_0 \leq |U|/(\frac{4\pi}{3} a^3)$. (Assumption (iii) of 7.7 is not required.)

Proof If $n > |U|/(\frac{4\pi}{3} a^3)$ and $q \in U\times\ldots\times U$ then $|q_i - q_j| \leq a$ for some i,j; $1 \leq i,j \leq n$, by geometry so that, since V is bounded below, $\sum_{j<k} V(|q_j - q_k|) = \infty$ for all $q \in U\times\ldots\times U$ and hence $Q_{U^n}(\beta) = 0$. □

For systems whose grand partition function is a polynomial, we have the following general properties.

7.9 Theorem In 7.4 (or more generally), suppose $\mathcal{Z}_U(z, \beta)$ is a polynomial in z of degree $N_0 \geq 1$, and $0 < |U| < \infty$. Then for $z \in R$, $z > 0$ we have

(i) $\beta P_U(z, \beta) > 0$

(ii) $0 < \frac{|U|}{N_0} \leq v_U(z, \beta) < \infty$ (see 7.8)

(iii) P_U and v_U are monotone increasing in z , and so, with abuse of notation, P_U is a function of v_U and β;

(iv) $\frac{\partial P_U}{\partial v_U}(v_U,\beta) = -\{\frac{\beta v_U^2}{\langle n\rangle}\langle n-\langle n\rangle\rangle^2\}^{-1} < 0$

(v) there is a neighborhood of the positive real axis on which $\mathcal{Q}_U$ has no zeros.

Proof We have

$$|U|\beta\ P_U(z,\beta) = \log(1 + a_1 z + \dots + a_N z^N)$$

and

$$\frac{1}{v_U(z,\beta)} = \frac{1}{|U|}\ \frac{a_1 z + a_2 z^2 + \dots + N a_N z^N}{1 + a_1 z + \dots + a_N z^N}$$

where $a_i \geq 0$ and the principal determination of the logarithm is understood.

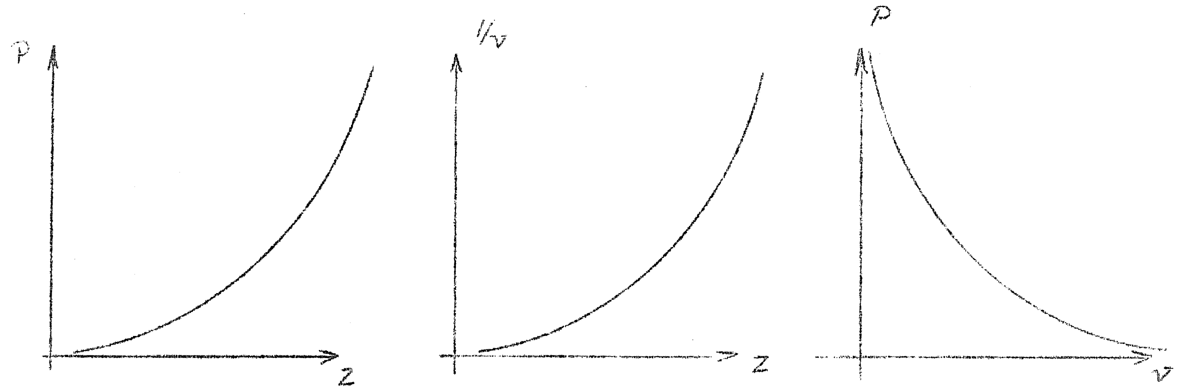
Hence (i) and (ii) are obvious. For (iii) we have, by direct differentiation,

$$|U|\beta\ \frac{\partial P}{\partial z}U(z,\beta) = \frac{a_1 + a_2 z + \dots + N a_N z^{N-1}}{1 + a_1 z + \dots + a_N z^N} = \frac{1}{z}\langle n\rangle$$

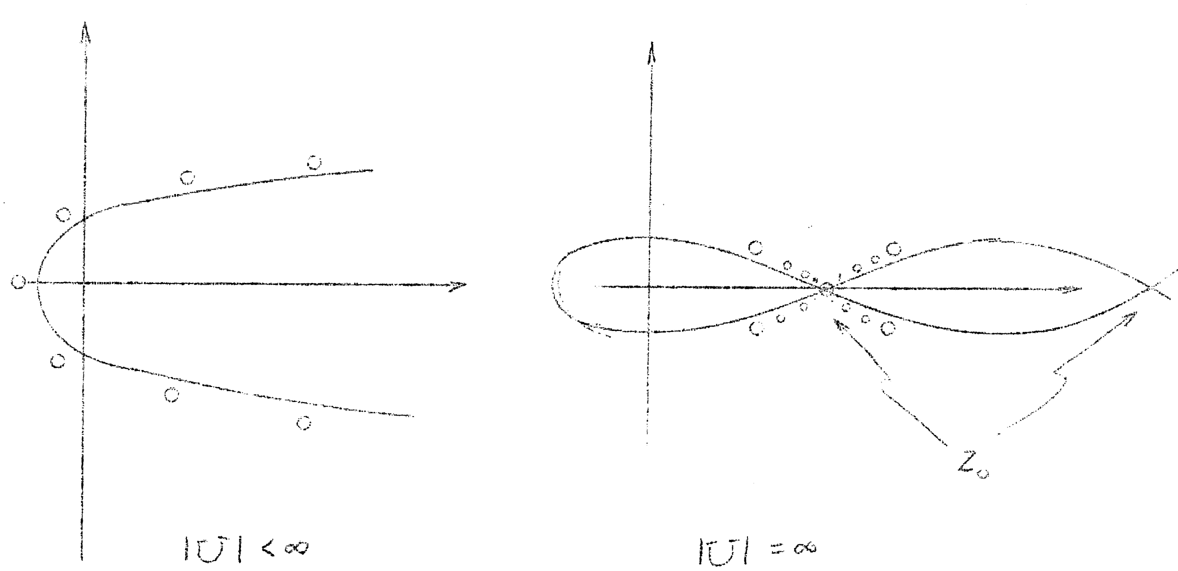
and $$-\frac{1}{v^2}\ \frac{\partial v}{\partial z} = \frac{\partial}{\partial z}\left(\frac{1}{v}\right) = \frac{1}{|U|\,z}\{\langle n^2\rangle - \langle n\rangle^2\} = \frac{1}{|U|\,z}\langle n-\langle n\rangle\rangle^2$$ (variance rule).

Thus (iii) is clear. For (iv), P is a differentiable function of v by the inverse function theorem, and $\frac{\partial P}{\partial v} = \frac{\partial P}{\partial z} / \frac{\partial v}{\partial z}$ which gives the stated formula. (v) is clear. □

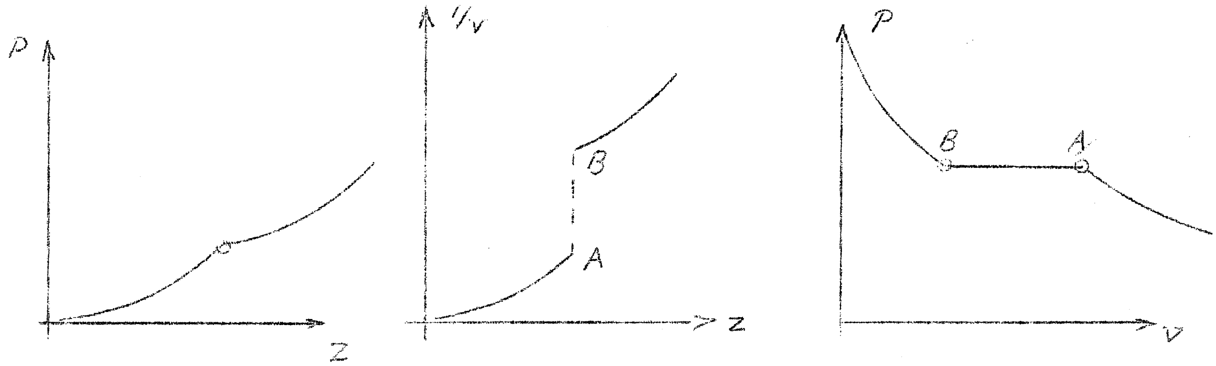
Thus in this case, all functions are analytic, indicating no phase transitions and we get behaviour typical of the gas phase:



However, as $|U| \longrightarrow \infty$ (thermodynamic limit), the situation is not so simple, for then the zeros of $\mathcal{Z}_U$ can converge on the real axis indicating a phase transition.



The Lee-Yang theorems guarantee that in the limit, P is continuous, but $^1/v$ may not be. Thus we have, qualitatively, the following:



The way U is allowed to expand is mildly restricted to the following.

7.10 <u>Definition.</u> Let $U_1, U_2, \ldots$ be an increasing sequence of subsets of R^3, and let $r_0 > 0$. We say that U_n are <u>r_0-regular</u> iff there are unions of non-overlapping cubes of side $\geq r_0$, T_n, W_n such that

(i) $T_n \subset U_n \subset W_n$

(ii) $|T_n| \longrightarrow \infty$ as $n \longrightarrow \infty$

(iii) $\dfrac{|W_n| - |T_n|}{|T_n|} \longrightarrow 0$ as $n \longrightarrow \infty$

(iv) diameter $T_n \longrightarrow \infty$. (so (iv) => (ii)).

Typical r_0-regular sets are expanding cubes, sphers, etc. More generally, if U_n is the magnification of a set with rectifiable boundary, they are r_0-regular.

The main theorems of Lee and Yang on the thermodynamic limit then may be stated as follows:

7.11 <u>Theorem (Lee-Yang)</u> In 7.4, suppose $V_{ij} = V$ is a Lee-Yang potential Then for z real, $z > 0$ and $\beta \in R$, $\beta > 0$, there is a non decreasing continuous function of z, $\beta P(z,\beta)$ such that for any r_0-regular sequence $U_n \subset R^3$ (r_0 is that of 7.7) we have

$$\operatorname*{limit}_{n \to \infty} \frac{1}{|U_n|} \log \mathcal{Z}_{U_n}(z,\beta) = \beta P(z,\beta)$$

where $\mathcal{Z}_{U_n}(z,\beta)$ is the grand partition function, defined in 7.3, 7.4.

The second theorem reduces the location of phase transitions to tracking the zero's of as V expands.

7.12 <u>Theorem (Lee-Yang)</u> Suppose the conditions of 7.11 hold and $\beta \in R$ is fixed. Let $S \subset C$ be an open, simply connected subset of C, the complex plane with $S \cap R^+ \neq \emptyset$ ($R^+ = \{x \in R: x \geq 0\}$) and $\mathcal{Z}_{U_n}(z,\beta)$ has no zeros in S for all n and $z \in S$. Then the convergence in 7.11 is uniform on compact subsets to an analytic (holomorphic) function $\beta P(z,\beta)$ in S. In addition,

$$\frac{1}{v(z,\beta)} = \underset{n \to \infty}{\text{limit}} \; \frac{1}{|U_n|} \; z \; \frac{\partial z}{\partial z} \{\log \mathcal{L}_{U_n}(z,\beta)\}$$

exists and is analytic in S, and is non decreasing for $z \in R$, $z > 0$.

Notice that S must be simply connected for 7.12 to make sense. See for example Ahlfors [1, p. 143]. (If S is simply connected, $f: S \to C$ analytic and nowhere zero there is an analytic function $\log f(z)$ such that $f(z) = \exp(\log f(z))$. Any two such functions differ by a multiple of $2\pi i$ (the determination).) We use the determination of the logarithm such that on the positive real axis we have the usual logarithm.

The proof of these theorems is somewhat involved so we prepare a number of lemmas: <u>The hypotheses of 7.11 will be assumed throughout.</u>

7.13 <u>Lemma</u>. Suppose $W_j \subset R^3$ is an increasing sequence of sets, $|W_j| < \infty$, $W'_j \subset W_j$ and $\underset{j \to \infty}{\text{limit}} \frac{|W_j| - |W'_j|}{|W'_j|} = 0$. Then we have

$$\underset{j \to \infty}{\text{limit}} \; \frac{1}{|W'_j|} [\log \mathcal{L}_{W_j}(z,\beta) - \log \mathcal{L}_{W'_j}(z,\beta)] = 0$$

for all $\beta \in R$, $z \in R$, $z \geq 0$.

<u>Proof</u>. Clearly, $Q^n_{W'_j}(z ,\beta) \leq Q^n_{W_j}(z ,\beta)$ as the integrands are positive and $W'_j \subset W_j$, and hence $W'_j x \ldots x W'_j \subset W_j x \ldots x W_j$. Thus, the polynomials $\mathcal{Q}$ satisfy $\mathcal{Q}_{W_j}(z ,\beta) \geq \mathcal{Q}_{W'_j}(z ,\beta)$.

We now claim that there is a constant $c > 0$ so that

$$\mathcal{Q}_{W_j}(z ,\beta) \leq \exp[z\, c(|W_j| - |W'_j|)]\, \mathcal{Q}_{W'_j}(z ,\beta)$$

which, together with the previous inequality will prove the lemma.

To do this, let $q_{n,\ell}(\beta)$ be the portion of $\frac{1}{n!}Q^n_{W_j}(\beta)$ corresponding to ℓ particles in $W_j \setminus W'_j$. More precisely,

$$q_{n,\ell}(\beta) = \frac{1}{\ell!(n-\ell)!\lambda^{3n}} \int_{G_\ell} \exp - \beta \sum_{i<j} V(|q_i - q_j|)dq$$

where $G_\ell \subset W_j\, x \ldots x\, W_j$ is defined by

$$G_\ell = \{q\colon\ q_1, \ldots q_\ell \in W_j \setminus W'_j\ ;\ q_{\ell+1}, \ldots, q_n \in W'_j\}$$

Hence by symmetry in q_j's we have

$$Q^n_{W_j}(\beta) = \sum_{\ell=0}^{m} q_{n,\ell}(\beta) \quad \text{where}$$

$M \leq \{n, |W_j| / \frac{4\pi}{3} a^3\}$ (see 7.8).

By Fubini's theorem, we have

$$\ell!(n-\ell)!\,\lambda^{3n} q_{n,\ell}(\beta) = \int_{W_j'^{\,n-\ell}} \int_{(W_j\setminus W_j')^{\ell}} \exp -\beta \sum_{i<j} V(|q_i-q_j|)dq_1\ldots dq_\ell)dq_{\ell+1}\ldots dq_n$$

$$\leq |W_j\setminus W_j'|^{\ell} \exp\{-\beta B\ell(\frac{r_0}{a})^3\} \int_{W_j'^{\,n-\ell}} \exp -\beta \sum_{i<j=\ell+1}^{n} V(|q_i-q_j|)dq_{\ell+1}\ldots dq_n$$

Since $V(r) \geq -B$, and for $q_i \in W_j\setminus W_j'$, $\sum_{j=i+1}^{n} V(|q_i-q_j|)$ contains at most $(\frac{r_0}{a})^3$ non zero terms, since we may assume $|q_i-q_j| > a$ as explained previously. Note that $|W_j\setminus W_j'| = |W_j| - |W_j'|$.

Thus, we have proven that

$$q_{n,\ell}(\beta) \leq \frac{1}{\ell!} \exp[-\beta\ell B(\frac{r_0}{a})^3](|W_j|-|W_j'|) \times \frac{1}{(n-\ell)!} Q^{n-\ell}_{W_j'}(\beta)$$

Therefore we have, since all sums are finite,

$$\mathcal{Q}_{W_j}(z,\beta) = \sum_{n=0}^{\infty} \frac{z^n}{n!} Q^n_{W_j}(\beta)$$

$$\leq \sum_{n\geq\ell} z^{\ell} \frac{[c(|W_j|-|W_j'|)]^{\ell}}{\ell!} \frac{z^{n-\ell}}{(n-\ell)!} Q^{n-\ell}_{W_j'}(\beta)$$

$$\leq \exp\ zc(|W_j|-|W_j'|)\ \mathcal{Q}_{W_j'}(z,\beta)$$

where $c = \exp[-\beta B(\frac{r_0}{a})^3] > 0$, and the lemma holds. $\square$

The inequality in the above proof will be required later.

7.14 Lemma. Suppose $U = \bigcup_{j=1}^{k} W_j$ where the distance between W_j and W_k; $j \neq k$ is at least r_0, (all substes of R^3) then if $|U| < \infty$,

$$\mathcal{Q}_U(z,\beta) = \prod_{j=1}^{k} \mathcal{Q}_{V_j}(z,\beta)$$

In particular, if V_j ~~ae~~ are all congruent by translation or rotation,

$$\mathcal{Q}_U(z,\beta) = [\mathcal{Q}_{V_j}(z,\beta)]^k$$

Proof. Clearly, $U^n = U \times \dots \times U = \cup\{W_{\sigma(1)}^{n_1} \times \dots \times W_{\sigma(k)}^{n_k} : \sigma$ is a permutation, $\Sigma\, n_i = n\}$ and by Fubini's theorem and the fact that $V(|q_j - q_k|)$ is either zero, or $q_j, q_k \in W_\ell$ for some ℓ we have $\frac{1}{n!} Q_U^n(\beta) = \Sigma\{Q_{W_i}^{n_1}(\beta)/ n_i! \dots n_k! : \Sigma n_j = n\}$ again by symmetry. The result for $\mathcal{Q}$ follows by the multinomial theorem for polynomials.□

We next prove the theorem for cubes.

7.15 Lemma. For $z > 0$, $\beta > 0$ there is a real number $\beta P(z,\beta)$ such that if W_j is a cube with side $2^j r_0$ then

$$\underset{j \to \infty}{\text{limit}} \frac{1}{|W_j|} \log \mathcal{Q}_{W_j}(z, \beta) = \beta P(z, \beta)$$

<u>Proof.</u> Partition off W_j into a union of 8^{j-i} cubes W_{jk}; $k = 1, \ldots, 8^{j-i}$ each with side $2^i r_0$, and let $W'_{jk} \subset W_{jk}$ be a concentric cube, a distance $r_0/2$ from the complement of W_{jk}. Then, by 7.14,

$$\mathcal{Q}_{W'_j}(z, \beta) = [\mathcal{Q}_{W'_{jk}}(z, \beta)]^{8^{j-i}}$$

where $W'_j = \bigcup_k W'_{jk}$.

From 7.13 then we conslude that

$$8^{j-i} \log \mathcal{Q}_{W'_{jk}}(z, \beta) \leq \log \mathcal{Q}_{W_j}(z, \beta)$$

$$\leq 8^{j-i} \log \mathcal{Q}_{W'_{jk}}(z, \beta) + z\, C(|W_j| - |W'_j|)$$

and, since $8^{j-i}|W_i| = |W_j|$, $\mathcal{Q}_{W_{jk}}(z, \beta) = \mathcal{Q}_{W_i}(z, \beta)$

$$\frac{1}{|W_i|}[\log \mathcal{Q}_{W'_{jk}}(z, \beta) - \log \mathcal{Q}_{W_{jk}}(z, \beta)]$$

$$\leq \frac{1}{|W_j|} \log \mathcal{Q}_{W_j}(z, \beta) - \frac{1}{|W_i|} \log \mathcal{Q}_{W_i}(z, \beta)$$

$$\leq \frac{1}{|W_i|}[\log \mathcal{Z}_{W'_{jk}}(z,\beta) - \log \mathcal{Z}_{W_{jk}}(z,\beta)] + zc(\frac{|W_j|-|W'_j|}{|W_j|}]$$

Now W_i satisfy the condition of 7.13 and

$$(|W_j| - |W'_j|) / |W_j| = \{1 - (1 - \frac{r_0}{2^i r_0})^3\} \longrightarrow 0 \quad \text{as} \quad i,j \longrightarrow \infty$$

so that $\frac{1}{|W_i|} \log \mathcal{Z}_{W_i}(z,\beta)$ forms a Cauchy sequence and, as R is complete, converge.□

7.16 Lemma. Suppose U_n is a union of $f(n)$ disjoint cubes with side $L \geq r_0$ and $f(n) \longrightarrow \infty$ as $n \longrightarrow \infty$. Then if diameter $U_n \longrightarrow \infty$,

$$\lim_{n \to \infty} \frac{1}{|U_n|} \log \mathcal{Z}_{U^n}(z,\beta) = \beta\, P(z,\beta)$$

where $\beta\, P(z,\beta)$ is given in 7.15.

Proof. From the argument of 7.15 we see that $\frac{1}{|U_n|} \log \mathcal{Z}_{U_n}(z,\beta)$ is a Cuachy sequence ($L \geq r_0$ is needed here). We leave the detailed verification of this to the reader. To see that it converges to the same limit as 7.15, select a subsequence U_n such that there are cubes $W_j \subset U_j$ where W_j is as in 7.15 and that $(|U_j|-|W_j|)/|U_j| \longrightarrow 0$. This can always be done. Then by the triangle inequality, we have

$$\left|\frac{1}{|U_j|}\log \mathcal{Z}_{U_j}(z,\beta) - \beta P(z,\beta)\right| \le \frac{1}{|U_j|}\left|\log \mathcal{Z}_{U_j}(z,\beta) - \log \mathcal{Z}_{W_j}(z,\beta)\right|$$

$$+ \left|\left(\frac{1}{|U_j|} - \frac{1}{|W_j|}\right)\log \mathcal{Z}_{W_j}(z,\beta)\right| + \frac{1}{|W_j|}\left|\log \mathcal{Z}_{W_j}(z,\beta) - \beta P(z,\beta)\right|$$

From the inequality of 7.13, the first term is dominated by $\frac{1}{|U_j|}$ $zc \times$ $(|U_j|-|W_j|) \longrightarrow 0$. Similarly, as $\frac{1}{W_j}$ $\log \mathcal{Z}_{W_j}(z,\beta)$ converges, the second term $\longrightarrow 0$. Hence a subsequence of $\frac{1}{|U_n|}\log \mathcal{Z}_{U_n}(z,\beta)$ converges to $\beta P(z,\beta)$. As the sequence is Cauchy, the whole sequence converges to $\beta P(z,\beta)$.□

Notice that it is not a priori obvious that by changing the cube size, or the speed at which they grow, will give the same limit as 7.15. This is the reason for 7.16.

<u>Proof of 7.11</u>. Find T_n and W_n as in 7.10 and note that 7.16 applies. The proof there shows that $\lim_{n\to\infty} \frac{1}{|U_n|}\log \mathcal{Z}_{U_n}(z,\beta) = \beta P(z,\beta)$ (the same triangle inequality).

Since $\log \mathcal{Z}_{U_n}(z,\beta)$ is non decreasing (see 7.9) and convergence is pointwise, $P(z,\beta)$ is also non decreasing.

To show it is continuous, fix $z_o > 0$ and find $\delta > 0$ so $z_0-\delta > 0$. Then from 7.9,

$$z \frac{\partial}{\partial z} [\frac{1}{|V|} \log \mathcal{Z}_U(z, \beta)] = \frac{1}{v_U(z, \beta)} \leq 1/(\frac{4\pi}{3} a^3)$$

a bound independent of U. Hence, on $(z_o - \delta, z_o + \delta)$, $\frac{1}{|U_n|} \mathcal{Z}_{U_n}(z, \beta)$ have iniformly bounded derivatives. Hence $P(z, \beta)$ is continuous at z_o. (If $f_n(x) \longrightarrow f(x)$ and f_n are differentiable with $f_n'(x) \leq M$ then f is continuous, for $|f_n(x) - f_n(y)| \leq |x-y|M$ which imples $|f(x) - f(y)| \leq |x - y|M$.) □

Next we turn to 7.12 and prepare some additional lemmas for this case.

7.17 Lemma (Vitali's theorem). Suppose f_n is a sequence of analytic functions on a connected open set $D \subset C$ and f_n are uniformly bounded on compact subsets of D. Then if there is a set $E \subset D$ so E has a limit point in D and for each $z_o \in E$, $f_n(z_o)$ converges then f_n converges uniformly on compact subsets to an analytic function f on D.

Recall that x is a limit point of E iff every neighborhood of ~~x~~ contains points of E other than x.

See, for example. Titchmarch [1, p. 169] or Whyburn [1, p. 87]. We shall also need the following fact from complex analysis (see Ahlfors [1]).

7.18 Lemma. Suppose $S \subset C$ is simply connected and open and $f\colon S \longrightarrow C$ is analytic with no zeros in S. Then $\log f(z)$ is analytic on S. Further, using the principal determination, if $|z| < 1$ then

$$\log(1 + z) = \sum_{i=1}^{\infty}(-1)^{i-1}\frac{z^i}{i}$$

uniformly and absolutly convergent on compact subsets of the unit disc.

To verify that $\frac{1}{U_n} \log \mathcal{Q}_{U_n}(z, \beta)$ are uniformly bounded on S (7.17) we first verigy the following:

7.19 Lemma. Let $D_\delta \subset S$ be a sphere centered at $\eta \in R$, $\eta > 0$ and of radius δ. Then $\frac{1}{U_n} \log \mathcal{Q}_{U_n}(z, \beta)$ are uniformly bounded on $D_{\delta/2}$. The hypotheses of 7.12 are assumed.

Proof. Let $z_1, \ldots, z_N$ be the zeros of the polynomial $\frac{1}{U_n} \log \mathcal{Q}_{U_n}(z, \beta)$, so that $N \leq |U| / \frac{4\pi}{3} a^3$. Let $z \in S$ and $y = z - \eta$, $y_i = z_i - \eta$ so that $\mathcal{Q}_{U_n}(z, \beta) = \pi(1 - \frac{z}{z_j}) = \pi(\frac{y_i - y}{y_i})(\frac{y_i}{i})$.

Then we have, using the principal determination,

$$\log \mathcal{Q}_{U_n}(z, \beta) = \sum_{i=1}^{N} \log\left(1 - \frac{y}{y_i}\right) + \sum_{i=1}^{N} \log\left(\frac{y_i}{z_i}\right)$$

(note that some care must be used in expressions like $\log z_1 z_2 = \log z_1 + \log z_2$ which generally hold modulo $2\pi i$) The last term is just $\log \mathcal{Z}_{U_n}(\eta,\beta)$, which by 7.11 when divided by U_n converges, so is uniformly bounded. We must show that $\frac{1}{|U_n|}\sum_{i=1}^{N} \log\left(1 - \frac{y}{y_i}\right)$ is informly bounded. But from 7.18, we have, since $\left|\frac{y}{y_i}\right| < 1$, (zeros lie outside D_δ),

$$\sum_{i=1}^{N} \log\left(1 - \frac{y}{y_i}\right) = -\sum_{i=1}^{N}\sum_{\ell=1}^{\infty}\left(\frac{y}{y_i}\right)^{\ell}\frac{1}{\ell}$$

$$= -\sum_{\ell=1}^{\infty}\sum_{i=1}^{N}\left(\frac{y}{y_i}\right)^{\ell}\frac{1}{\ell}$$

which we can rearrange, by absolute convergence, on $D_{\delta/2}$, say. But

$$\left|\sum_{i=1}^{N}\left(\frac{y}{y_i}\right)^{\ell}\frac{1}{\ell}\right| \leq \frac{|y|^{\ell}}{\ell}\sum_{i=1}^{N}\frac{1}{|y_i|^{\ell}}$$

$$\leq \frac{|y|^{\ell}}{\ell}\frac{N}{\delta^{\ell}} \leq \left(\frac{y}{\delta}\right)^{\ell} N$$

so that $\left|\sum_{i=1}^{N}\log\left(1 - \frac{y}{y_i}\right)\right| \leq N/\left(1 - \left|\frac{y}{\delta}\right|\right)$

But $N/|U_n|$ is uniformly bounded, and if $|y| \leq \delta/2$ we have the result.□

<u>Proof of 7.12.</u> By 7.19 and 7.17, $\frac{1}{U_n}\log \mathcal{Z}_{U_n}(z,\beta)$ converges uniformly to an analytic function on $D_{\delta/2}$. For any compact set $C \subset S$ we cover it with a finite number of discs and we may assume one disc intersects R^+. Proceeding inductively, by 7.19 we see that $\frac{1}{U_n}\log \mathcal{Z}_{U_n}(z,\beta)$

is uniformly bounded.

The statements for v follow at once, for if f_n are analytic and converge uniformly to f then f_n' converge uniformly to f'.□

From this proof, we also have:

7.20 Corollary. In 7.12, we have for $z \in S$,

$$\frac{1}{v(z,\beta)} = \beta z \frac{\partial P}{\partial z}(z,\beta)$$

Notice that a phase transition does not necessarily result from an analytic function being non continuable across a singularily, but from the fact that the various analytic portions may not match up. (Compare the Mayer theory).

Lee and Yang [2] went on and applied this theory to the Ising Model (a Hamiltonian for a set of spins fixed on a lattice in a magnetic field; see Huang [1, Ch. 16]), and showed that the zeros of $\mathcal{Q}$ all lie on the unit circle and converge to 1. Thus there are two phases with transition at $z = 1$. Everywhere else on R, P and $1/v$ are analytic.

Even before Yang and Lee, Van Hove [1] asserted the existence in the thermodynamic limit of the free energy and pressure in the grand canonical and canonical ensembles (see § 9). However, it was pointed out by Van Kampen that the proof was fallacious. See Fisher [1, footnote 2], where the correct proof also appears. In § 9 we shall generalize some of the above results (due mainly to Ruelle and Fisher).

§8 Stability

This section is concerned with the concept of "stability" and the sufficient conditions for it introduced by Ruelle. It is important for the thermodynamic limit; and more specifically, for generalizations of the theorems of Lee and Yang (§7), discussed in the next section.

These notions are important for quantum mechanical systems as well as classical ones. Much of this section follows Ruelle [2]. We begin with the definition of stability.

8.1. Definition. Consider a grand canonical ensemble consisting of phase spaces M^n and Hamiltonian functions H^n: $M^n \longrightarrow R \cup \{\infty\}$; not necessarily smooth for $n = 0,1,2,\ldots.$ The system is called stable iff there is a constant $B > 0$ such that

$$H^n(m) = \geq -Bn$$

for all $n = 0,1,\ldots$ and $m \in M^n$.

The first important remark is the following.

8.2. Theorem. Consider a grand canonical ensemble with H^n: $T^*M^n \longrightarrow R$, $H^n = T^n + V^n$ where T^n is the kinetic energy obtained from a Riemannian metric and V^n is a function on M^n. Then H^n is stable iff there is a constant $B > 0$ so

$$V^n(q) \geq -Bn$$

for all $q \in M^n$, $n = 0,1,\ldots$

Further, suppose $U^n \subset M^n$ lies in a compact set and $|U^n| \leq K^n$ for a constant $K > 0$. Then for each $\beta > 0$, $z > 0$, $\sum_{n=0}^{\infty} \frac{z^n}{n!} \int_{T^*U^n} e^{-\beta H^n} d\mu$ converges. (We assume the dimension of M^n is kn.)

<u>Proof</u>. If H^n is stable then $V^n(q) > - Bn$ by taking $p = 0$ ie. consider the zero section. The converse is clear as T^n is positive.

For the second part, it suffices to prove the result in a coordinate chart. By Fubini's theorem $\int_{T^*U^n} e^{-\beta H^n} d\mu = \int_{U^n}(\int_{R^n} e^{-\beta T^n} d\mu)(e^{-\beta V^n} dq)$. The first factor was evaluated in §7 in terms of the eigenvalues of g, and is bounded by $1/\lambda^n$ for a constant $\lambda > 0$. (The details are left to the reader.) The result now follows from:

8.3 <u>Lemma</u>. Suppose $a > 0$ and $\beta > 0$; then

$$\sum_{n=0}^{\infty} \frac{a^n}{n!} e^{+\beta n} \quad \text{converges } (< \infty)$$

and

$$\sum_{n=0}^{\infty} \frac{a^n}{n!} e^{\beta n^2} \quad \text{diverges } (= \infty).$$

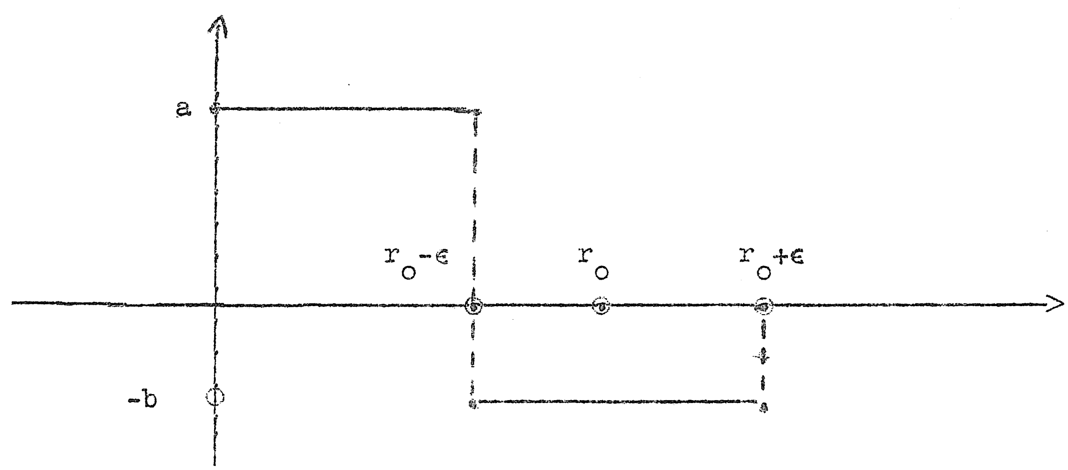
Proof. The first series is $\exp(ae^{\beta}) < \infty$. The second diverges by the ratio test, for example.□

Next we consider conditons equivalent to stability for a grand canonical ensemble arising from two body forces which have no hard cores. To issustrate what is going on, first sonsider an example

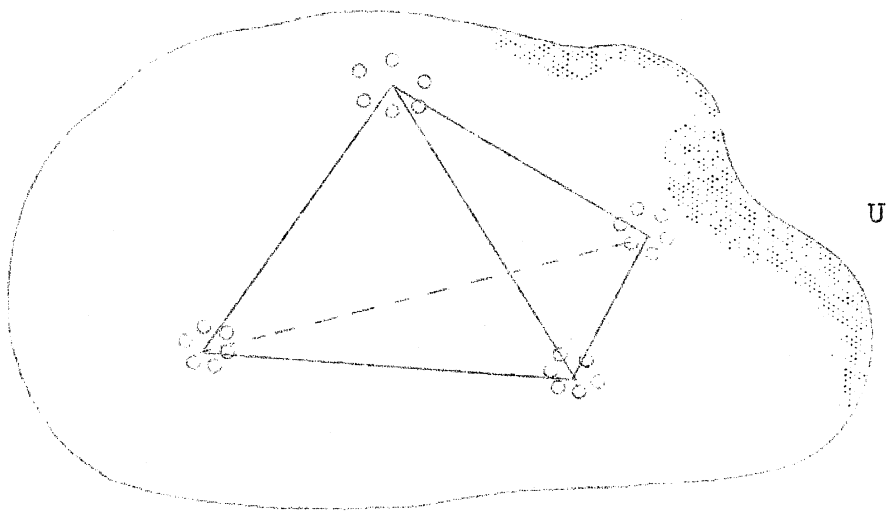
8.4 Example (An unstable system). Consider a two body potential $V: R^{+} \longrightarrow R$;

$$V(r) = \begin{cases} a & \text{if } r \le r_o - \epsilon \\ -b & \text{if } r_o - \epsilon < r < r_o + \epsilon \\ 0 & \text{if } r \ge r_o + \epsilon \end{cases}$$

where a, b, r_o, ϵ are positive real numbers, $r_o - \epsilon > \epsilon$, and $3b > a$



The corresponding grand canonical ensemble on $R^{3n} \times R^{3n}$ is unstable (7.4). To see this, consider points $x_1, \ldots x_{4s}$ with s points in the $\epsilon/2$-neighborhood of a tetrahedron in R^3.



It is clear that

$$\sum_{i,j=1}^{4s} V(|x_i - x_j|) = 4sa + 4\ s(s-1)a - 12\ bs^2$$

$$= 4s^2 a - 12\ s^2 b < 0$$

Hence $$\sum_{i<j} V(|x_i - x_j|) = -\frac{4s}{2} V(0) + \sum_{i,j=1}^{4s} V(|x_i - x_j|)$$

$$= -2sa + (4a - 12b)s^2$$

and so, by the lemma, the grand partition function diverges, for a region $U^n = U \times \ldots \times U$ and U containing a tetrahedron as above.

Next we prepare the following:

8.5 Definition. Let X be a topological space and $f: x \longrightarrow R \cup \{\infty\}$ a mapping. We say f is upper semi-continuous iff for every $\epsilon > 0$ and $x \in X$ there is a neighborhood U of x such that $y \in U$ implies $f(y) \leq f(x) + \epsilon$. Lower semi-continuous is defined analogously.

Obviously, $f: X \longrightarrow R$ is continuous iff it is both upper and lower semi-continuous. Also if $g: Y \longrightarrow X$ is continuous and $f: X \longrightarrow R$ upper semi-continuous then $f \circ g$ is upper semi-continuous, (other hypotheses when $X = R$ generally fail).

Notice that the next theorem does not apply to systems with hare cores; even though in that case the grand partition function is a polynomial, as we saw in 7.8, they can be unstable.

8.6 Theorem (Ruelle). Suppose $V: R^m \longrightarrow R \cup \{\infty\}$ is an upper semi-continuous function with $V(0) < \infty$ and $V(-x) = V(x)$. (Let $m > 0$ be a fixed integer.) Then the following are equivalent

(i) for any $x_1, \ldots, x_n \in R^m$,

$$\sum_{i,j=1}^{n} V(x_i - x_j) \geq 0$$

(ii) there is a $B \in R$, $B > 0$ so

$$\sum_{i<j=1}^{n} V(x_i - x_j) \geq -Bn \qquad \text{(stability)}$$

(iii) for any $U \subset R^m$ measureable with $|U| < \infty$ (Lebesque measure) and all $z > 0$, $\beta > 0$

$$\mathcal{L}_U(z,\beta) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \int_{U \times \ldots \times U} \exp[-\beta \sum_{i<j} V(x_i - x_j)] dx_1 \ldots dx_n$$

converges

(iv) for any $U \subset R^m$, $|U| < \infty$, and some $z > 0$, $\beta > 0$, $\mathcal{L}_U(z,\beta)$ in (iii) converges.

Upper semi-continuity is not required for the equivalence of (i) and (ii) nor the implication (i) => (iii) and (iv).

Proof. Since $\sum_{i,j=1}^{n} V(x_i - x_j) = n\, V(0) + 2 \sum_{i<j} V(x_i - x_j)$ it is clear that (i) and (ii) are equivalent. That (ii) implies (iii) follows since

$$\mathcal{L}_U(z,\beta) \leq \sum_{n=0}^{\infty} \frac{z^n}{n!} |U|^n \exp(\beta B n)$$

which converges. (This was also noted in 8.2.)

As (iii) implies (iv), it remains to show that (iv) implies (i). Suppose (i) is false. Then there are $x_1, \ldots, x_n$ in R^m so $\sum_{i,j=1}^{n} V(x_i - x_j) = -2\epsilon < 0$.

By upper semi-continuity (see remarks after 8.5) there are neighborhoods W_i of x_i so that $y_i \in W_i$ implies $\sum_{i,j=1}^{n} V(y_i - y_j) < -\epsilon$. Let $W = \bigcup_{i=1}^{n} W_i$ and $W' = W_1 \times \ldots \times W_n$ (s-times). Then for $q \in W'$,

$$\begin{aligned}\sum_{i<j=1}^{ns} V(q_i - q_j) &= -\frac{ns}{2} V(0) + \sum_{i,j=1}^{n} V(q_i - q_j) \\ &\leq -\frac{ns}{2} V(0) - \frac{(ns)^2}{2} \epsilon\end{aligned}$$

so that

$$\sum_{n=0}^{\infty} \frac{z^n}{n!} \int_{W \times \ldots \times W} \exp[-\beta \sum_{i<j} V(q_i - q_j)] dq$$

$$\geq \sum_{n=0}^{\infty} \frac{z^n}{n!} \int_{W'} \exp[-\beta \sum_{i<j} V(q_i - q_j)] \, dq$$

diverges for any $z > 0$, $\beta > 0$ by 8.3.□

Clearly (using translation invariance) (iv) can be replaced by

(iv)' for any sphere $U \subset R^n$, and some $z > 0$, $\beta > 0$, $\mathcal{L}_U(z,\beta)$.

converges.

There is a convenient and quite general criterion for stability which we next sonsider, due to Fisher and Ruelle [1]. First, recall the following

8.7 <u>Definition</u>. Let $\mathcal{H}$ be a complex Hilbert space (2.3) and $T: \mathcal{H} \longrightarrow \mathcal{H}$ a linear map. Then T is called <u>positive</u> iff for each $x \in \mathcal{H}$, $\langle x, Tx \rangle \in R^+$ (ie. is real and nonnegative).

A complex $n \times n$ matrix is called <u>positive</u> iff for any orthonormal basis, the corresponding linear transformation is positive (in $C^n = C \times \ldots \times C$) (The condition is easily seen to be independent of the basis.)

8.8 <u>Lemma</u>. If $T: \mathcal{H} \longrightarrow \mathcal{H}$ is positive then it is symmetric (self-adjoint, Hermitian).

An nxn matrix A_{ij} is postive iff for all $z_1,\ldots,z_n \in C$,

$$\sum_{i,j=1}^{n} \bar{z}_i A_{ij} z_i \geq 0$$

Proof. Obviously $\langle x,Tx\rangle = \langle Tx,x\rangle$ as it is real. By the polarization identity ($\langle x,y\rangle = \frac{1}{4}\{\langle x+y,x+y\rangle - \langle x-y,x-y\rangle + i\langle x+iy,x+iy\rangle - i\langle x-iy,x-iy\rangle\}$) we have the result. See Halmos [3,p.13]. The second part is clear as the expression is just $\langle z,Az\rangle$ where $z = (z_1,\ldots,z_n)$ with respect to the standard basis.□

The notation in the Fisher-Ruelle theorem below is somewhat cumbersome, so we deal with it separately.

8.9 Definition. Let $m > 0$ be an integer and $M^n = R^{mn} \times R^{mn}$; $H^n\colon M^n \longrightarrow R \cup \{\infty\}$ with $n_1 + n_2 + \ldots + n_\mu = n$, be a grand canonical ensemble (§7). We say it arises from μ-spácies of particles iff $H^n = T^n + V^n$ where T^n is obtained from a Riemannian metric (kinetic energy ≥ 0) and where $V^n\colon R^{nm} \longrightarrow R \cup \{\infty\}$ is of the following form:

$$V^n(q) = \sum_{\alpha=1}^{\mu} \sum_{i(\alpha)<j(\alpha)} \varphi_{\alpha\alpha}(q_{i(\alpha)} - q_{j(\alpha)})$$

$$+ \sum_{\alpha<\beta=1}^{\mu} \sum_{i(\alpha)} \sum_{j(\beta)} \varphi_{\alpha\beta}(q_{i(\alpha)} - q_{j(\beta)})$$

where $\varphi_{\alpha\beta}: R^m \longrightarrow R \cup \{\infty\}$ for $\alpha,\beta = 1,\ldots,\mu$ and with $q \in R^{mn}$ denoted $q = (q_{i(1)},\ldots,q_{1(n_1)},\ldots,q_{\mu(1)},\ldots,q_{\mu(n_\mu)})$.

The theorem then gives a sufficient condition for such a system to be stable. (Notice that Fisher-Ruelle add a redundant hypothesis $\varphi_{\alpha\beta}(q) = \varphi_{\beta\alpha}(-q)$ which follows from those of 8.10, 8.8.)

8.10 Theorem. In 8.9 suppose $\varphi_{\alpha\beta} = \varphi^{(1)}_{\alpha\beta} + \varphi^{(2)}_{\alpha\beta}$ where $\varphi^{(1)}_{\alpha\beta}(q) \geq 0$, and possibly a positive many body potential is added to H^n. If we can write, for $\hat{\varphi}_{\alpha\beta}: R^m \longrightarrow R$,

$$\varphi_{\alpha\beta}(x) = \int_{R^m} e^{ixy}\, \hat{\varphi}_{\alpha\beta}(y)dy$$

where $\hat{\varphi}_{\alpha\beta}(y)$ is a postive $\mu \times \mu$ matrix, and $\varphi_{\alpha\beta}(0) < \infty$. Then the system is stable. In fact

$$V^n(q) \geq -\frac{1}{2}\sum_{\alpha=1}^{\mu} n_\alpha\, \varphi^{(2)}_{\alpha\alpha}(0)$$

Proof. Clearly

$$V^n(q) \geq \sum_{\alpha=1}^{\mu} \sum_{i(\alpha)<j(\alpha)} \varphi^{(2)}_{\alpha\alpha}(q_{i(\alpha)} - q_{j(\alpha)})$$

$$+ \sum_{\alpha<\beta} \sum_{i(\alpha)} \sum_{j(\beta)} \varphi^{(2)}_{\alpha\beta}(q_{i(\alpha)} - q_{j(\beta)})$$

$$= \frac{1}{2} W^n(q) - \frac{1}{2}\sum_{\alpha=1}^{\mu} n_\alpha\, \varphi^{(2)}_{\alpha\alpha}(0)$$

where $W^n(q) = \sum\limits_{\alpha,\beta,i,j} \varphi^{(2)}_{\alpha\beta}(q_{i(\alpha)} - q_{j(\beta)})$

$$= \Sigma \int \exp i\, y(q_{i(\alpha)} - q_{j(\beta)})\, \varphi_{\alpha\beta}(y)dy$$

≥ 0, from the lemma.□

8.11 Corollary. Suppose $H^n(q,p) = \frac{1}{2} g(p,p) + \sum_{i<j} V\,(q_i - q_j) + V'(q)$ where $V'(q) \geq 0$, and

$$V(x) = \int e^{iyx}\, \hat{V}(y)dy,\ \hat{V}(y) \geq 0,$$

with $V(0) < \infty$. Then the system is stable.

The property of V in 8.11 is called positive type. More generally, we have:

8.12 Theorem (Bochner). A function $f\colon R^m \longrightarrow R$ is of positive type (that is, for any $x_1,\ldots,x_n \in R^m$ $z_1,\ldots z_n \in C$, $\sum\limits_{i,j=1}^{n} \bar{z}_i\, f(x_i - x_j) z_j \geq 0$) iff there is a measure μ_f on R^m, $\mu_f(R^m) < \infty$ so

$$f(x) = \int e^{iyx}\, d\mu_f(y)$$

Moreover, f is integrable, resp. continuous, in L^2 iff $\mu_f = \hat{f}\,\mu_o$ where μ_o is Lebesque measure and $\hat{f}$ is continuous; resp. integrable, in L^2.

For the proof, see Bochner [1], for example, or Schwartz [2].

This makes the proof of 8.11 trivial:

Alternative proof of 8.11. We have

$$\sum_{i,j=1}^{n} V(x_i - x_j) \geq 0$$

using $z_i = 1$. Hence, since $V(x) = V(-x)$, 8.6 applies.□

The corollary 8.11 has wide application as we shall see later. First, we note a few more consequences:

8.13 Corollary. Suppose, in 8.9, there are constants q_α so $\varphi_{\alpha\beta}(x) = q_\alpha q_\beta X(x)$ where $X\colon R^m \longrightarrow R \cup \{\infty\}$ is of positive type, $X(0) < \infty$. Then the system is stable.

Proof. $\hat{X}(y) \geq 0$ so that $\sum_{\alpha,\beta} z_\alpha q_\alpha q_\beta \bar{z}_\beta \hat{X}(y) = |\sum_\alpha (z_\alpha q_\alpha)|^2 \hat{X}(y) \geq 0$ so 8.10 applies.□

The corollary does not hold for the coulomb potential, but will if we smear the particles slightly.

8.14 Corollary. In 8.9, suppose $\rho_\alpha\colon R^m \longrightarrow R$ are continuous with compact support (for example), $X\colon R^m \longrightarrow R \cup \{\infty\}$ is of positive type and

$$\varphi_{\alpha\beta}(x) = \int_{R^m \times R^m} \rho_\alpha(y)\rho_\beta(y')\, X\,(x+y-y')dydy'$$

with $\varphi_{\alpha\alpha}(0) < \infty$. Then the system is stable.

Proof. From the convolution theorem (see Yosida [1], for instance), and $\hat{\rho}_\alpha(p) = \overline{\hat{\rho}_\alpha(-p)}$

$$\hat{\varphi}_{\alpha\beta}(y) = \hat{\rho}_\alpha(p)\, \overline{\hat{\rho}_\beta(p)}\, \hat{X}(p)$$

(modulo some positive factor), so the same calculation as 8.13 shows $\hat{\varphi}_{\alpha\beta}(y)$ is a postive matrix.□

Examples of such X are the coulomb ($X(x) = 1/|x|$) and Yukawa ($X(x) = e^{-K|x|}/|x|$) potentials. We leave the computation of these Fourier transforms to the reader.

For an amusing direct proof of 8.14 for the Coulomb case (going back to Onsager in 1939), see Fisher-Ruelle [1, p. 263].

Fisher and Ruelle also discuss the quantum mechanical cse, and show, under general circumstances that we have catastrophy when the stability condition is not satisfied. (like the divergence in 8.3).

Stability in the pure coulomb case in quantum mechanics has recently been established for fermions by Dyson and Lenard. See Dyson [1]. The proof of this important fact is, however quite difficult.

In naclear physics the problem of stability was also discussed (See Blatt and Wersskopf [1, Chapter III]) but the main emphasis was on giving criteria under which exchange forces (not considered here) could produce stability in a system which would otherwise be unstable.

The remainder of this section is devoted to more useful tests for stability. We begin with hard cores (see 7.8 and compare 8.6).

8.15 Theorem. Consider, on R^m, a potential $V: R^m \longrightarrow R \cup \{\infty\}$, $V(x) = V(-x)$ and a hard core of radius $a > o$ (see §7). Suppose there is a constant $D > 0$ so $x_1, \ldots, x_n \in R^m$, $|x_i - x_j| > a$ implies $\sum_{i=1}^{n} V(x_i) \geq -D$. Then the system with V as two body force is stable (7.4).

Proof. Obviously $\infty = \sum_{i<j} V(x_i - x_j) \geq -\frac{n}{2} D$ if $|x_i - x_j| \leq a$ for some i,j. But if $|x_i - x_j| > a$ for all $i,j = 1,\ldots,n$ then

$$\sum_{i<j=1}^{n} V(x_i - x_j) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j \neq i} V(x_i - x_j) \geq -\frac{1}{2} nD,$$

since $|(x_i - x_j) - (x_i - x_k)| \geq a.\Box$

8.16 Corollary. The Lee-Yang two body potential (7.7) is stable.

Proof. If $V(r) \geq -B$ and $|x_i - x_j| > a$,

$$\sum_{i=1}^{n} V(x_i) \geq -B\,(r_o/a)^3$$

by geometry. □

For example, nuclear forces are often of the Lee-Yang type. More delicate conditions ensuring 8.15 are the following

8.17 Corollary (Ruelle-Penrose). Suppose $V\colon R \longrightarrow R$ is a two body potential with hard core of radius $a > 0$. Suppose V is continuously differentiable on (a,∞) and

(i) $\lim_{r \to \infty} V(r) = 0$

(ii) $\int_a^\infty r^m \left|\frac{dV}{dr}(r)\right| dr < \infty$

Then the corresponding grand canonical ensemble on R^m is stable (7.4).

Proof. We assume the reader is familiar with the Riemann-Stieltjes integral (Apostol [1, Ch. 9]). Assume $m = 3$ for simplicity. Let $x_1, \ldots, x_n \in R^3$, $|x_i - x_j| > a$ and $F_i(r)$ denote the number of x_j, $j \neq i$ with $|x_i - x_j| \leq r$, so that

$$\sum_{j \neq i} V(|x_i - x_j|) = \int_a^\infty V(r)\,dF_i(r)$$

$$= V(r)F_i(r) \Big|_a^\infty - \int_a^\infty F_i(r) \frac{dV}{dr}(r)\,dr$$

by the integration by parts formula. The first term is zero by (i). But

$$\left| \int_a^\infty F_i(r) \frac{dV}{dr}\,dr \right| \leq \int_a^\infty \left(\frac{r}{a}\right)^3 \left|\frac{dV}{dr}\right| dr < \infty$$

Hence from 8.15 we have the result (c.f. proof of 8.15).□

The next theorem exploits the method of 8.11.

8.18 Theorem. Suppose $V: R \longrightarrow R \cup \{\infty\}$ satisfies

(i) V is bounded below

(ii) there are positive real numbers $a_1, a_2, c_1, c_2, n_1, n_2$ (with $n_1, n_2 \geq m$ in R^m) such that

$$V(r) > c_1 r^{-n_1} \quad \text{if} \quad r \leq a_1$$

$$|V(r)| < c_2 r^{-n_2} \quad \text{if} \quad r \geq a_2$$

Then we can write $V = V_1 + V_2$ where $V_1 \geq 0$ and V_2 is continuous of positive type. In particular, the grand canonical ensemble with V as two body potential is stable (8.15).

For the proof we shall employ the following

8.19 Lemma. Let $\alpha > 0$ and $F\colon R^m \longrightarrow R$, $F(x) = (|x|^2 + \alpha^2)^{-n/2}$ for $n \geq m$. Let $G(\xi) = (1 + |\xi|^2)^{-n/2}$ so that $\alpha^{m-n}\,\hat{G}(\alpha p) = \hat{F}(p)$ where $\hat{F}$ is the Fourier transform of F. Then G is positive and decreasing.

The proof is a technicality we omit. Hints are outlined in Ruelle [2], and the Fourier transform is calculated explicitly in Schwartz [2, p. 116].

Proof of 8.18. Multiplying V by a constant if necessary, we may assume $a_1 < 1 < a_2$. Let $n = \min(n_1, n_2)$ so that

$$V(r) - c_1 r^{-n} \geq 0 \quad \text{for} \quad r < a_1$$

and

$$|V(r)| < c_2 r^{-n} \qquad \text{for} \quad r > a_2$$

and hence

$$|V(r) - c_1 r^{-n}| \leq (c_1 + c_2) r^{-n} \quad \text{for} \quad r > a_2.$$

Since V is bounded below ((i)) there are positive constants β, c_3 so

$$V(r) - c_1 r^{-n} > -c_3 (r^2 + \beta^2)^{-n/2}$$

for all r, as is easily seen.

From 8.19, there is α, $0 < \alpha < \beta$ so that

$$\hat{V}_2(p) = c_1\alpha^{m-n}\,\hat{G}(\alpha p) - c_3\beta^{m-n}\,\hat{G}(\beta p) \geq 0$$

for all $p \in R^m$.

Let $V_1(r) = V(r) - V_2(r)$

$$= V(r) - c_1(r^2 + \alpha^2)^{-n/2} - c_3(r^2 + \beta^2)^{-n/2}$$

$$\geq c_1[r^{-n} - (r^2 + \alpha^2)^{-n/2}] \geq 0$$

which completes the proof.□

8.20 Corollary. In 8.18, consider in R^m,

$$V(r) = \frac{A}{r^{n_1}} - \frac{B}{r^{n_2}}\ ;\ n_1 > n_2 \geq m,\ A,B > 0$$

Then the corresponding system is stable. (Lenard-Jones potentials.)

Proof. (i) of 8.18 is clear as $V(r)$ is continuous, $V(r) \longrightarrow 0$ as $r \longrightarrow \infty$ and $V(r) \longrightarrow \infty$ as $r \longrightarrow 0$. For (ii), $V(r) = (A-Br^{n_1-n_2})/r^{n_1}$ and $A - Br^{n_1-n_2} > C_1$ for r sufficiently small. The other inequality is similar.□

§9. The Thermodynamic Limit of the Thermodynamic Functions.

Here we discuss the thermodynamic limit in the three ensembles; microcanonical (§3), canonical (§7) and grand canonical (§7) with respect to the thermodynamic functions. Correlation functions on the other hand, will be considered in §10.

We concentrate mainly on the basic ingredients of the subject and the thermodynamic limit in the spirit of Yang and Lee (§7). For the equivalence of the formalisms we have a few remarks, but refer to Van der Linden [1] for the main theorems.

Our assumption on the potentials is that of strong tempering. Fisher [1] assumes only weak tempering for the limit in the canonical ensemble. Whether this is sufficient for all three ensembles is not known at present.

The reader should compare the above mentioned treatments with ours, as the proofs and methods are quite different. The only new results here are a simplification and extension of Van der Linden [1, §3 and 5], and some differentiability criteria for the various functions. We also give precise treatment of differentiation with respect to volumes, a subject that is not usually carefully discussed in this context.

We begin with the definitions for the microcanonical ensemble.

9.1 Definition. Consider phase spaces $M^n \subset R^{n\nu} \times R^{n\nu}$ and Hamiltonians $H^n : M^n \longrightarrow R$ (we assume that all singular parts such as hard cores and collisions have been removed from M^n) $n = 0,1,2,\cdots$. For $D \subset M^n$, (measurable), let

$$\Sigma\,(n,D,e) = \mu\{m \in D : H^n(m) < ne\}/n!$$

where μ is the measure on M^n. (Think of e as the energy per particle)
If $D = U \times \cdots \times U \times R^{n\nu} \cap M^n$, we write $\Sigma\,(n,U,e)$ for $\Sigma\,(n,D,e)$.

The <u>microcanonical</u> <u>entropy</u> is defined by:

$$S^{(m)}(n,U,e) = \frac{1}{n} \log \Sigma\ (n,U,e)$$

In the following we shall assume that H^n is symmetric under permutation of $(q_1,\dots,q_n)$.

Notice that $\Sigma\ (n,U,e)$ is a positive increasing function of e and D (in the sense of inclusion). It is also worthwhile noticing that if H^n is a stable system (8.1), then $\Sigma\ (n,U,e) = 0$ for all $e < -B$. (The $n!$ and $1/n$ in the above are essential for the theory.) (See the delicacy of 9.23, 9.22).

9.2 <u>Theorem</u>. Suppose $H^n : M^n \longrightarrow R$ is a (stable) system as in 9.1. Then we have, in the sense of Riemam-Stieltjes integration, for each $\beta > 0$, and e the "variable",

$$\int_{-\infty}^{\infty} \exp(-ne\beta)d\ \Sigma\ (n,U,e) = \frac{1}{n!} \int_D \exp(-\beta H)d\mu$$

(Note that this makes sense without Σ being differentiable, as it is monotone).

<u>Proof</u>. This is a special case of the following fact from measure theory (See Halmos [1,p. 67,80]): If μ is a measure on a space Y and $u : Y \longrightarrow R$ with $g(t) = \mu\{x \in Y : u(x) < t\}$ and $f:R \longrightarrow R$ is continuous with compact support, then

$$\int_Y f\circ u\ d\mu = \int_{-\infty}^{\infty} f\ dg$$

This easily gives the theorem.$\square$

This theorem states the connection of the microcanonical with the canonical ensemble. The right side is just Q_D.

It is a famous theorem of Lebesque that a monotone function on R is differentiable almost everywhere. However, we shall require a stronger condition to ensure that this derivative will coincide with the distributional derivative (see 9.16 below). The result is as follows:

9.3 Theorem. Suppose, in 9.1, that D is open and that (for n fixed),

(i) $\{m \in D : H(m) < ne\}$ lies in a compact set,

and (ii) H is smooth (C^∞) on D

Then Σ is differentiable almost everywhere in e. More specifically, if ne is a regular value of H (that is, $dH(m) \neq 0$ if $m \in H^{-1}(ne)$) then Σ is C^∞ in a neighborhood of e.

Proof. For this proof we assume a knowledge of calculus on manifolds. See Abraham [1, ch II] for example. By Sard's theorem, it suffices to prove the last statement. Now for each $m \in H^{-1}(ne)$ there exists, by the implicit function theorem, a local chart about m in which H becomes linear. We may suppose this local chart is a rectangle for example. The set in (i) is covered by a finite number of these charts. But

$$\frac{1}{h}\{\Sigma(e+h) - \Sigma(e)\} = \frac{1}{h}\mu\{m \in D : e \leq H(m) < e+h\} \quad (\text{if } h > 0),$$

has a limit on each of the charts and hence on the finite union. The theorem readily follows. $\square$

In case
$$D = U \times \cdots \times U \times R^{n\nu} \cap M^n, \text{ and}$$
$$H^n(q,p) = \sum_{i=1}^{n} p_i^2/2m + V_n(q)$$
notice that (i) holds if U is open and bounded. Also,

$$dH(q,p) = 0 \text{ iff } p = 0 \text{ and } dV_n(q) = 0$$

(equilibrium point).

Next we wish to define $\partial\Sigma / \partial U$. For this, we require:

9.4 Lemma. In 9.1, suppose ne is a regular value of H^n. Then there is a unique positive function $f : R^\nu \longrightarrow R$, continuous almost everywhere such that for each $U \subset R^{\nu}$ measurable,

$$\Sigma\ (n,U,e) = \int_U f(x)\,dx.$$

(in fact, f is smooth on an open dense set). dx denotes Lebesque measure.

Proof. We know that f exists by the Radon-Nikodym theorem. The point here is continuity almost everywhere. However by Fubini's theorem we see that f is obtained by integration of a characteristic function, that of the set $\{m \cdot H(m) < ne\}$ which is open. In fact,

$$f(x) = \frac{1}{n!} \int x_e d_\mu \quad \text{where } n\text{-}1 \text{ copies of } U\ U \times \cdots \times U \times R^{n\nu}$$

appear and x_e is the characteristic function of $\{m : H(m) < ne\}$. Since the discontinuities of this characteristic function x_e, are a closed set of measure zero, as $H^{-1}(ne)$ is a submanifold (implicit function theorem), we have the result.□

9.5 Definition. In 9.1, 9.4, suppose U is an open submanifold whose boundary ∂U is a submanifold and is compact, (or a submanifold almost everywhere, using charts), so that ∂U inherits a natural measure. Then define

$$\frac{\partial\Sigma}{\partial U}\,(n,U,e) = \int_{\partial U} f\,d\mu_o,$$

where μ_o is the measure on ∂U. Notice that this definition defines $\partial\Sigma/\partial U$ uniquely except on sets where f may fail to be continuous on a set of positive measure on ∂U. (This collection of surfaces has, in a precise sense, which we won't discuss, measure zero).

In the case of regular shapes such as spheres or cubes, $\partial\Sigma/\partial U$ may be defined by differentiating with respect to the linear dimensions of U as follows:

9.6 Theorem. In 9.1, suppose U is the interior of a sphere (resp. cube) centered at the origin in R^ν and $\lambda U = \{\lambda x : x \in U\}$ for $\lambda \in R$. Then if f (in 9.5) is continuous at almost all points of ∂U (which will happen for almost all spheres in view of 9.5),

$$\frac{\partial \Sigma}{\partial U}(n,U,e) = \frac{d}{d\lambda}\Sigma(n,\lambda U,e) \quad \text{at} \quad \lambda = 1.$$

Proof. Evidently we must show

$$\int_{\partial U} f \, d\mu_o = \lim_{h \to 0} \frac{1}{h}\left(\int_{(1+h)U} f \, d\mu - \int_U f \, d\mu\right)$$

Since f is bounded we may assume it is continuous, and hence uniformly continuous. Then if f_o is defined to be constant on the radii, and equal to f on ∂U we have by Fubini's theorem,

$$\int_{\partial U} f \, d\mu_o = \frac{1}{h}\int_{(1+h)U} f_o \, d\mu - \int_U f_o \, d\mu)$$

$$= \frac{1}{h}\{\int_{(1+h)U} (f-f_o) \, d\mu - \int_U (f-f_o) \, d\mu\}.$$

For h sufficiently small, say $h > 0$, $|f-f_o| < \epsilon$ on the radii of $(1+h)U/U$, so the above is $\leq \frac{1}{h}\epsilon\, |(1+h)U/U| \leq \epsilon K$ for a constant K (K is the area of a sphere in R^ν). Hence the assertion. □

The above derivative is with respect to a scale, not the volume. For the latter, a factor $(\nu\lambda^{\nu-1})^{-1}$ is inserted. We define then

$$\frac{\partial \Sigma}{\partial |U|} = \frac{1}{\nu}\frac{\partial \Sigma}{\partial U}.$$

Further quantities of interest in the microcanonical ensemble are as follows:

9.7 Definition. In 9.1, define

(i) (Microcanonical temperature $^{-1}$)

$$\beta^{(m)}(n,U,e) = \partial s^{(m)}(n,U,e)/\partial e$$

(under the conditions of 9.3, say; a.e).

(ii) (Microcanonical pressure)

$$p^{(m)}(n,U,e) = \frac{\partial\Sigma}{\partial|U|}(n,U,e)/\ \Sigma(n,U,e)$$

with $\partial\Sigma/\partial U$ defined in 9.5 for suitable sets U, and $\partial\Sigma/\partial|U| = (\partial\Sigma/\partial U)/\nu$ as above.

Here one should think of $|U|/n$ as being constant (density) so that 9.7(ii) essentially agrees with the definition of pressure in §6.

In the case of the canonical ensemble we have a similar set of definitions:

9.8 Definition In 9.1, define (see 7.1)

(i) (canonical partition function)

$$Q(n,U,\beta) = \frac{1}{n!}\int_{U\times\cdots\times U\times R^{n\nu}} \exp(-\beta H_n)\, d\mu$$

(ii) (canonical free energy)

$$f^{(c)}(n,U,\beta) = \frac{1}{n}\log Q(n,U,\beta)$$

(iii) (canonical internal energy)

$$e^{(c)}(n,U,\beta) = \frac{1}{n}< H_n > = \frac{1}{n}\int \exp(-\beta H_n)H d\mu/\int \exp(-\beta H_n)d\mu = -\ \partial f^{(c)}(n,U,\beta)/\partial\beta.$$

(iv) (canonical pressure)

$$p^{(c)}(n,U,\beta) = n\frac{\partial f^{(c)}}{\partial|U|}(n,U,\beta),$$

$$(\text{ie:} \frac{\partial Q}{\partial|U|}(n,U,\beta)\ /\ Q(n,U\beta);$$

Note Q is a measure on U, but $f^{(c)}$ is not).

One has a result for Q similar to that of 9.6.

The connection with the microcanonical ensemble is:

9.9 Theorem. In the above, we have

(i) $Q(n,U,\beta) = \int_{-\infty}^{\infty} \exp(-\beta ne)\, d\,\Sigma\,(n,U,e)$

(ii) $e^{(c)}(n,U,\beta) = \int_{-\infty}^{\infty}\exp(-\beta ne)\, e\, d\,\Sigma(n,U,e)/Q$

(iii) $p^{(c)}(n,U,\beta) = \int_{-\infty}^{\infty} \exp(-\beta ne) \cdot n\, d\{\frac{\partial \Sigma}{\partial |U|}(n,U,e)\}/Q$

Proof. This follows readily from 9.2 and the definitions. □

Thus, in a certain sense, the canonical quantities are means of the microcanonical ones in a probability distribution. In the limit one obtains equalities $(n \longrightarrow \infty)$ using the central limit theorem. That is a far from obvious fact first proved in van der Linden [1] which will not be proved here.

Finally we come to the grand canonical ensemble: (Here it is slightly more convenient to use μ than z)

9.10 Definition. In 9.1, let (c.f. 7.4)

(i) (grand canonical partition function)

$$\mathcal{Q}(U,\beta,\mu) = \sum_{n=0}^{\infty} e^{\mu n} Q\,(n,U,\beta)$$

(say the system is stable;§8)

(ii) (grand canonical pressure)

$$p^{(g)}(U,\beta,\mu) = \frac{1}{|U|} \log \mathcal{Q}(U,\beta,\mu)$$

(iii) (grand canonical specific volume)

$$1/U^{(g)}(U,\beta,\mu) = - \partial p^{(g)}(U,\beta,\mu)/\partial \mu$$

(iv) (grand canonical internal energy)

$$e^{(g)}(U,\beta,\mu) = - U^{(g)}(U,\beta,\mu)\frac{\partial}{\partial \beta} p^{(g)}(U,\beta,\mu).$$

The connection with the canonical ensemble is (c.f. 9.9)

9.11 Theorem. In the above, we have

(i) $1/\,U^{(g)}(U,\beta,\mu) = \langle n \rangle/|U|$

$$= \frac{1}{|U|} \sum_{n=0}^{\infty} e^{\mu n}\, n\, Q(n,U,\beta)$$

(ii) $$e^{(g)}(U,\beta,\mu) = \frac{1}{<n>} \sum_{n=0}^{\infty} \frac{e^{\mu n}}{n!} \int e^{-\beta H_n} H_n \, d\mu/2$$
$$= \frac{1}{<n>} <H_n>.$$

Again this follows at once from the definitions. The preceding definitions set up the three thermodynamic analogies using the microcanonical, canonical and grand canonical ensembles. These can be expected to coincide only in the thermodynamic limit. Thus, one of the fundamental problems of classical statistical mechanics is the following. (For a recent review see Mazur [1].)

9.12 Asymptotic Problem of Classical Statistical Mechanics.

The problem is, roughly, to prove the following statements under reasonable hypotheses:

I. Microcanonical Thermodynamic Limit.

(1) For each $v \in R$, $v > 0$ show that

$$\lim_{n \to \infty} s^{(m)}(n,U_n,e) = s^{(m)}(e,v)$$

exists, for "any" $U_n \subset R^\nu$ with $|U_n|/n \to v$ and that $s^{(m)}(e,v)$ so defined is concave in e and v.

(2) Show that, as in (1), almost everwhere, we have

$$\lim_{n \to \infty} \beta^{(m)}(n,U_n,e) = \partial s^{(m)}(e,v)/\partial e$$

(3) Show that also,

$$\lim_{n \to \infty} p^{(m)}(n,U_n,e) = \partial s^{(m)}(e,v)/\partial v \quad \text{a.e.}$$

II. Canonical Thermodynamic Limit

(4) As in (1), show that

$$\lim_{n \to \infty} f^{(c)}(n,U_n,\beta) = f^{(c)}(\beta,v)$$

exists and is convex in β, concave in v.

(5) Show that

$$\underset{n \to \infty}{\text{limit}}\ e^{(c)}(n, U_n, \beta) = \partial f^{(c)}(\beta, v)/\partial\beta \quad \text{a.e.}$$

(6) $$\underset{n \to \infty}{\text{limit}}\ p^{(c)}(n, U_n, \beta) = \partial f^{(c)}(\beta, v)/\partial v \quad \text{a.e.}$$

III. <u>Grand Canonical Thermodynamic Limit.</u>

(7) Show that, for suitable $U \subset R^{\nu}$,

$$\underset{|U| \to \infty}{\text{limit}}\ p^{(g)}(U, \beta, \mu) = p^{(g)}(\beta, \mu)$$

exists and is convex in β, μ

(8) Show that

$$\underset{|U| \to \infty}{\text{limit}}\ 1/v^{(g)}(U, \beta, \mu) = \partial q^{(g)}(\beta, \mu)/\partial\mu$$

(9) Show

$$\underset{|U| \to \infty}{\text{limit}}\ e^{(g)}(U, \beta, \mu) = - v^{(g)}(\beta, \mu)\partial p^{(g)}(\beta, \mu)/\partial\beta .$$

IV. <u>Equivalence of the three Ensembles.</u>

(10) Show that the microcanonical and canonical ensembles are equivalent, in the sense that if $\beta(v,e)$ is defined implicitly by

$$e^{(c)}(\beta, v) = e$$

then

$$s^{(m)}(e, v) = f^{(c)}(\beta(e, v), v) + e\,\beta(e, v)$$

(11) Show that the canonical and grand canonical ensembles are equivalent. That is if $\mu(\beta, v)$ is defined implicitly by

$$v^{(g)}(\beta, \mu) = v, \quad \text{then}$$

$$f^{(c)}(\beta, v) = p^{(g)}(\beta, \mu(\beta, v))v + \mu(\beta, v).$$

This definition is what we mean by the thermodynamic limit for the classical thermodynamic functions. For n finite notice that we have an honest classical Hamiltonion system with a flow. At present, there seem to be no theorems showing that the thermodynamic limit is an honest classical Hamiltonion system with an infinite number of degrees of freedom. This passage should be roughly, the change from ordinary to partial differential equations.

The rest of this section is devoted to the proof of (1) - (9) above. For (10) and (11) we refer to van der Linden [1] (the idea is as stated in 9.9 and 9.11.)

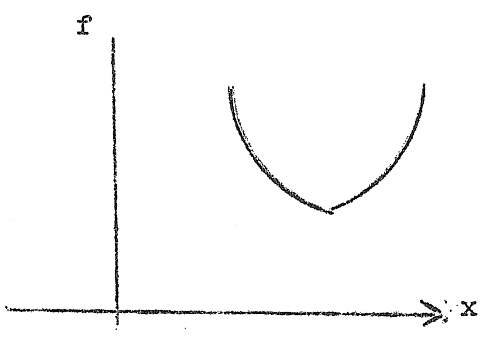
In this section we will not attempt generalizing the second theorem of Lee and Yang (7.12). This will be discussed in § 10.

First, we recall a few basic facts about convex functions:

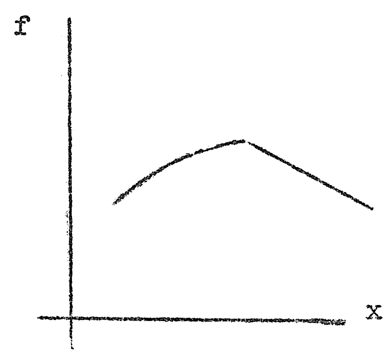
9.13 Definition. Suppose $f: R (\text{or } R^+) \longrightarrow R$. We say f is convex iff for all $x,y \in R$, $0 < \alpha < 1$ we have

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y).$$

Similarly f is concave iff $-f$ is convex.



f convex



f concave

There are some basic and standard facts about convex functions we now list:

9.14 Lemma. (i) if f is measurable then f is convex iff

$$f(\tfrac{1}{2}(x+y)) \leq \tfrac{1}{2}(f(x) + f(y));$$

(ii) If f is convex then f is continuous and is differentiable almost everywhere. The derivative is monotone increasing;

(iii) If f is twice differentiable, f is convex iff $d^2f/dx^2 > 0$.

See, for example Hardy-Littlewood-Polya [1]. From (ii), observe that df/dx has only a countable number of jump discontinuities (Apostol [1,p 162]).

To attack the above problems, it is now standard to use the following theorem due to Griffiths [1].

9.15 Lemma. Let $f_n : R(\text{or}R^+) \longrightarrow R$ be a sequence of convex functions with $f_n \longrightarrow f$ almost everywhere. Then

(i) f is convex (obvious),

(ii) with the exception of a countable set of points x

$$\underset{n \longrightarrow \infty}{\text{limit}} \frac{df_n}{dx}(x) = \frac{df}{dx}(x).$$

However, this (with "countable" replaced by "of zero measure")is a special case of a more powerful result which we now give; it assumes an elementary knowledge of distribution theory.

9.16 Lemma. Suppose f_n and f are locally bounded functions and the distributional derivatives df_n/dx and df/dx are locally bounded functions. Then if $f_n \longrightarrow f$ a.e., we have $df_n/dx \longrightarrow df/dx$ almost everywhere.

Proof. Since $f_n \longrightarrow f$ a.e., $f_n \longrightarrow f$ in the sense of distributions. Hence $df_n/dx \longrightarrow df/dx$ in the sense of distributions as the operation of differentiation is continuous. But if $g_n \longrightarrow g$ in the sense of distributions and g_n, g are locally bounded functions, then $g_n \longrightarrow g$ a.e. To see this, suppose there is an $\epsilon > 0$ and a set A of positive measure such that $|g_n - g| \geq \epsilon$ on A, for an infinity of n. By the regularity of Lebesque measure, there is an open set U so $\mu(U) = \mu(A)$. $U \supset A$. Choose a smooth function φ one on an open subset of U, and zero outside U. Then we have

$$\int |g_n\varphi - g\varphi| \, d\mu \geq \epsilon \, \mu\,(A)$$

So g_n cannot converge to g. □

From this lemma we may dispose of many of the above problems easily

First we show that (1) implies (2) and (3)

9.17 Theorem. In the microcanonical ensemble, 9.1, suppose that H is smooth. Then

(i) if $U_n \subset R^\nu$ and a.e.,

$$\lim_{n \to \infty} s^{(m)}(n,U_n,e) \to s^{(m)}(e,v) \text{ and}$$

$s(m)$ is convex in e then

$$\lim_{n \to \infty} \beta^{(m)}(n,U_n,e) = \partial s^{(m)}(e,v)/\partial e \text{ a.e.}$$

(ii) if for every sequence of spheres (or cubes) with $|U_n|/n = v$, and some e, (a regular value of H_n),

$$\lim_{n \to \infty} s^{(m)}(n,U_n,e) = s^{(m)}(e,v)$$

and $s^{(m)}$ is convex in v then

$$\lim_{n \to \infty} p^{(m)}(n,U_n,e) = \partial s^{(m)}(e,v)/\partial v \,.$$

Proof. From 9.3, $s^{(m)}(n,U_n,e)$ is locally smooth a.e. and from 9.14, $s^{(m)}(e,v)$ has a locally bounded derivative. Hence 9.16 gives (i). To prove (ii), we have $n\,\partial\Sigma/\partial|U| = \partial\Sigma/\partial v$ where $v = |\lambda U|/n$ and this holds locally for almost all v by 9.6. Hence 9.16 again gives (ii). □

Next we show that part of (4) implies (4),(5) and (6);

9.18 Theorem. In 9.8, $f^{(c)}(n,U,\beta)$ is convex in β

(i) If $\lim_{n \to \infty} f^{(c)}(n,U_n,\beta) = f^{(c)}(\beta,v)$ for all β then $f^{(c)}(\beta,v)$ is convex in β and

$$\lim_{n \to \infty} e^{(c)}(n, U_n, \beta) = \partial f^{(c)}(\beta, v)/\partial \beta \quad \text{a.e.}$$

(ii) If, for all spheres (or cubes) with

$$|U|/n = v, \qquad \lim_{n \to \infty} f^{(c)}(n, U_n, \beta) = f^{(c)}(\beta, v)$$

(β fixed), then

$$\lim_{n \to \infty} p^{(c)}(n, Un, \beta) = \partial f^{(c)}(\beta, v)/\partial v$$

almost everywhere (v).

Proof. For (i), it is sufficient to check

$$\partial^2 f^{(c)}(n, U, \beta)/\partial \beta^2 \geq 0 \text{ in view of } 9.14,\ 9.15$$

However we have

$$\frac{\partial^2 f^{(c)}}{\partial \beta^2}(n, U, \beta) = < H_n >^2 - < H_n^2 > = <(H_n - < H_n >)^2 > \geq 0$$

by a simple calculation. (Means in the canonical ensemble).

(ii) follows in a way similar to 9.17 (ii). □

Next we show that part of (7) implies (8) and (9).

9.19 Theorem. In 9.10, suppose the system is stable and for some sequence $U_n \subset R^\nu$ and for almost all $\beta, \mu \in R$,

$$\lim_{n \to \infty} p^{(g)}(U_n, \beta, \mu) = p^{(g)}(\beta, \mu)$$

exists. Then

(i) $p^{(g)}(U_n, \beta, \mu)$ and $p^{(g)}(\beta, \mu)$ are convex in β, μ

(ii) $\lim_{n \to \infty} 1/v^{(g)}(U_n, \beta, \mu) = \partial p^{(g)}(\beta, \mu) \partial \mu$, a.e.

and (iii) $\lim_{n \to \infty} e^{(g)}(U_n, \beta, \mu) = - v^{(g)}(\beta, \mu) \dfrac{\partial p^{(g)}}{\partial \beta}(\beta, \mu)$, a.e.

Proof. By 9.14, 9.15 it is sufficient to check (i). Also we may assume the sum over n is finite, again by 9.15. We have

$$\frac{\partial^2}{\partial\mu^2} \log 2(U,\beta,\mu) = <(n - < n >)^2 > \geq 0$$

so that $p^{(g)}$ is convex in μ, and similarly

$$\frac{\partial^2}{\partial\beta^2} \log 2\ (U,\beta,\mu) = <(H_n - < H_n>)^2> \geq 0$$

so that $p^{(g)}$ is convex in β.□

The remaining problems (1), (4), (7) will be treated simultaneously. There is a further basic condition required besides stability which we now introduce. By the standing convention, box potentials are excluded from the regions (See 7.4 for example). (It seems to us that the ideas of "free volume" (see Fisher [1]) here are quite unnecessary; see §7).

9.20 <u>Definition</u>. Consider $H^n : M^n \subset R^{n\nu} \times R^{n\nu} \longrightarrow R$

$$H^n(q,p) = \sum_{i=1}^{n} p_i^2/2m + V_n(q).$$ We say the system is <u>strongly tempered</u> iff

(i) V_n is invariant under translation in $R^{n\nu}$ and under permutation of the coordinates $q_i \in R^\nu$

(ii) there exists $r_o > 0$ such that for all n,m,

$$V_{n+m}(x_1,\cdots,x_n,y_1,\cdots,y_m) \leq V_n(x_1,\cdots,x_n) + V_m(y_1,\cdots,y_m)$$

whenever $|x_i - y_j| \geq r_o$, for all $i=1,\cdots,n$ and $j=1,\cdots,m$. Here $y_j, x_i \in R^\nu$.

For example, in the case of two body forces:

9.21 <u>Lemma</u>. If, in 9.20, $V_n(q) = \sum_{i<j=1}^{\nu} V(q_i - q_j)$

for $V : R^\nu \longrightarrow R \cup \{\infty\}$, $V(q) = V(-q)$ and $V(x) \leq 0$ if $|x| \geq r_o$, then the system is strongly tempered.

Proof. Clearly (i) holds. Observe that

$$V_{n+m}(x_1,\ldots,x_n,y_1,\ldots,y_m) = V_n(x_1,\ldots,x_n) + V_m(g,\ldots,y_m) + \sum_{i,j=1}^{n,m} V(x_i-y_j)$$
$$\leq V_n(x_1,\ldots,x_n) + V_m(y_1,\ldots,y_m). \square$$

It is useful to retain 9.20 in the generality of many body forces, as these actually occur. (chemistry for example). Notice that hard cores are not excluded by 9.20 or stability.

Weak tempering requires only that there is ϵ, W so that for each n,m,

$$V_{n+m}(x_1,\ldots,x_n,y_1,\ldots,y_m) \leq V_n(x_1,\ldots,x_n) + V_m(y_1,\ldots,y_m) + nmWr^{-(\nu+\epsilon)}$$

if r is sufficiently large, where $|x_i-y_i| > r$. (r depends on n,m).

The above lemma shows that for two body forces, one demands

$$V(x) \leq D|x|^{-(\nu+\epsilon)} \quad \text{if} \quad |x| \geq r_o.$$

Next we require two sets of inequalities derived from stability and strong tempering respectively for the thermodynamic limit. (These play the role of 7.13 and 7.14).

9.22 Lemma. Suppose

$$H^n : M^n \subset R^{n\nu} \times R^{n\nu} \longrightarrow R \quad H^n(q,p) = \sum_{i=1}^{n} p_i^2/2m + V_n(q)$$

is a stable system. That is, $V_n(q) \geq -nB$ for a constant B. Let δ_m be the measure of the unit ball in R^m, so $\delta_m = \pi^{m/2}/\Gamma(1+m/2)$ where Γ is the usual gamma function ($\Gamma(1+n) = n!$ if n is an integer).

Then we have: (c.f. 9.1, $\hat{9}$.8 and 9.10).

(i) $\Sigma(n,U,e) \leq \dfrac{|U|^n}{n!} \delta_{n\nu} [2m(e+B)n]^{n\nu/2}$

(i)' $s^{(m)}(n,U,e)$ is bounded above ($|U|/n, e$ fixed)

(ii) $Q(n,U,\beta) \leq \dfrac{|U|^n}{n!\lambda^{n\nu}} \exp(n\beta B)$

(ii)' $f^{(c)}(n,U,\beta)$ is bounded above $(|U|/n,\beta$ fixed)

(iii) $2(U,\beta,\mu) \leq \exp\{\frac{|U|}{\lambda^{\nu}} \exp(\beta B-\mu)\}$

(iii)' $p^{(g)}(U,\beta,\mu)$ is bounded above. $(\beta,\mu$ fixed).

Proof. We have

$$n!\,\Sigma(n,U,e) = \mu\{(q,p) \in U\times\cdots\times U\times R^{n\nu}\cap M^n : H^n(q,p) < ne\} \leq \mu(A\times B)$$

where $A = U\times\cdots\times U$ and $B = \{p\in R^{n\nu} : \Sigma p_i^2/2m \leq (e+B)n\}$

by subset inclusion. Hence (i) is clear. To prove (i)', we use (i), and $\log\Gamma(1+\alpha) \geq x\log x - x$ (Stirlings formula) to deduce

$$\begin{aligned} s^{(m)}(n,U,e) &= \frac{1}{n}\log\Sigma(n,U,e) \\ &\leq \frac{1}{n}\cdot n[\log|U| + \frac{\nu}{2}[1+\log(4\pi m(e+B)/\nu)] \\ &\quad - n\log n + n \\ &= \log(|U|/n) + \frac{\nu}{2}[1+\log(4\pi m(e+B)/\nu)]+1 \end{aligned}$$

(No terms may be dropped, as this is fairly delicate). (ii) and (iii) were proven in §8 and (ii)' and (iii)' follow directly, using Stirlings formula for (ii)'. □

Next the inequalities using strong tempering.

9.23 Lemma. Suppose $H^n : M^n \subset R^{n\nu}\times R^{n\nu} \longrightarrow R : H^n(q,p) = \sum_{i=1}^{n} p_i^2/2m + V_n(q)$ is a strongly tempered system. Suppose $U', U'' \subset U \subset R^{\nu}$ and the distance between U' and U'' is at least r_o of 9.20. Then we have (c.f., 9.1, 9.8 and 9.10)

(i) $\Sigma(n,U,e) \geq \sum_{m=o}^{n} \Sigma(m,U',\frac{n}{m}e')\,\Sigma(n-m,U'',\frac{n}{n-m}(e-e'))$ for all $e' \leq e$.

(i)' $s^{(m)}(n+m,U,e)\cdot n+m \geq s^{(m)}(n,U',e)\cdot n + s^{(m)}(m,U'',e)\cdot m$

(ii) $Q(n,U,\beta) \geq \sum_{m=o}^{n} Q(m,U',\beta)\, Q(n-m,U'',\beta)$ for all $\beta > 0$.

(ii)' $(n+m)\, f^{(c)}(n+m,U,\beta) \geq n\, f^{(c)}(n,U',\beta) + m\, f^{(c)}(m,U'',\beta)$

(iii) $\mathcal{Z}(U,\beta,\mu) \geq \mathcal{Z}(U',\beta,\mu)\,\mathcal{Z}(U'',\beta,\mu)$

(iii)' $|U|\, p^{(g)}(U,\beta,\mu) \geq |U'|\, p^{(g)}(U',\beta,\mu) + |U''|\, p^{(g)}(U'',\beta,\mu)$.

<u>Proof</u>. To prove (i), we have

$$\Sigma(n,U,e) = \frac{1}{n!}\mu\{(q,p) \in U \times\cdots\times U \times R^{n\nu} \cap M^n : \Sigma\, p_i^2/2m + V_n(q) < ne\}$$

$$\geq \frac{1}{n!}\mu\{(q,p) \in U' \cup U'' \times\cdots\times U' \cup U'' \times R^{n\nu} \cap M^n :$$

$$\Sigma\, p_i^2/2m + V_n(q) < ne\}$$

by set inclusion.

Let $$A_m = \{q \in U' \cup U'' \times\cdots\times U' \cup U'' \cap M^n :$$

exactly m components of q lie in $U'\}$. Clearly A_m are disjoint measurable sets, so that

$$\Sigma\,(n,U,e) \geq \frac{1}{n!}\sum_{m=o}^{n} \mu\{(q,p) \in A_m \times R^{n\nu} : \Sigma\, p_i^2/2m + V_n(q) < e\}$$

$$\geq \frac{1}{n!}\sum_{m=o}^{n} \frac{n!}{m!(n-m)!}\mu\{(q_m,q_{n-m},p_m,p_{n-m}) \in$$

$$U' \times\cdots\times U' \times U'' \times\cdots\times U'' \times R^{m\nu}\times R^{(n-m)\nu} : \sum_{i=1}^{m} p_i^2/2m + V_m(q_m) < ne'$$

and $$\sum_{i=n-m+1}^{n} p_i^2/2m + V_{n-m}(q_{n-m}) < n(e-e')\}$$

again by set inclusion and permutation symmetry. ((i) of 9.20). The set above is a cartesian product so that, rearranging the factorials, we have (1). To prove (i)', choose one term of (i) to get

$$\log \Sigma\,(n+m,U,e) \geq \log \Sigma\,(n,U',\frac{n+me'}{m}) + \log \Sigma\,(m,U',\frac{n+m}{m}(e-e'))$$

Now choose $e' = \frac{n}{n+m}e$ so that this is exactly the result (i)!

The proof of (ii) and (iii) follow by a similar argument using the U, U', U" in the cartesian product. (See 7.17 also). Taking logarithms gives (ii)' and (iii)' . □

The inequalities (i)',(ii)',(iii)' will now be used to obtain the thermodynamic limit. The argument is based on the elementary theory of subadditive functions which we now review.

9.24 <u>Definition.</u> A sequence of real numbers $a_1, a_2, \cdots$ is called <u>subadditive</u> iff $a_{n+m} \le a_n + a_m$ for all integers n,m. Similarly a mapping $f : R(\text{or } R^+) \longrightarrow R$ is subadditive iff $f(x+y) \le f(x) + f(y)$ for all $x,y \in R$ (resp. R^+).

The notions of subadditivity and convexity are quite different and should not be confused. A basic fact about subadditive sequences we need is:

9.25 <u>Lemma.</u> (Polya - Szegö [1, pt. I, exercise 98]) Let $a_1, a_2, \cdots$ be a subadditive sequence and suppose $\{a_n/n\}$ is bounded below. Then a_n/n converges i.e. is a Cauchy sequence in R.

<u>Proof.</u> Let $a = \liminf \{a_n/n\} = \lim\limits_{n \to \infty} \inf\limits_{k \ge n} \{a_k/k\}$ so $a \in R$ as a_n/n is bounded below. Given $\epsilon > 0$ choose N so $n \ge N$ implies $a_n/n > a - \epsilon$, and find some $m_o \ge N$ so $a_{m_o}/m_o < a + \epsilon/2$. Let $K_\epsilon = \max\{a_1, \cdots, a_{m_o}\}$ and choose $N \ge N$ so $n \ge N$, implies $K_\epsilon/n < \epsilon/2$. We claim that $n \ge N$, implies

$$a - \epsilon < a_n/n < a + \epsilon.$$

To see this, find an integer r so $rm_o \le n \le (r+1)m_o$. Then by subadditivity,

$$a_n/n \le \frac{1}{n}(arm_o + a_{n-rm_o}) \le \frac{1}{n}\{ra_m + K_\epsilon\} \le a_m/m + K_\epsilon/n \le a + \epsilon/2 + \epsilon/2 = a + \epsilon. \;\square$$

Evidently if a_n/n is not bounded below, then $a_n/n \longrightarrow -\infty$.

Although we shall not need it, it is interesting to note a generalization of the above. Namely, if $a_{n_1,\cdots,n_r}$ is a multisequence subadditive in each index, and $a_{n_1,\cdots,n_r}$ is bounded below, then it converges. This can be used below to obtain the thermodynamic limit of rectangles (rather than cubes) directly.

As in §7, to obtain the thermodynamic limit, we require some regularity conditions on the regions. The condition here is slightly stronger to compensate for slightly weaker inequalities. (The reader should compare these with the definitions of Fisher[1]).

9.26 <u>Definition</u>. Let $r_o \in R$, $r_o > 0$, and $U_j \subset R^\nu$, $j=1,2,\cdots$. We say that U_j are <u>strongly r_o-regular</u> iff for any $L \geq r_o$, there are unions of cubes of side L ; say T_j, W_j such that

(i) $$T_j \subset U_j \subset W_j \; ; \; j = 1,2,\cdots$$

and (ii) $$(|W_j|-|T_j|) / |T_j| \longrightarrow 0$$

Notice that from this it follows that

$$|W_j| / |U_j| \quad \text{and} \quad |T_j| / |U_j| \longrightarrow 1 \quad \text{as} \quad j \longrightarrow \infty.$$

Then the main theorem on the thermodynamic limit may be stated as follows:

9.27 <u>Theorem</u>. Suppose $H^n : M^n \subset R^{n\nu} \times R^{n\nu} \longrightarrow R$,

$$H^n(q,p) = \sum_{i-1}^{n} p_i^2/2m + V_n(q)$$

is a stable, strongly tempered system. Then

(i) there exists a function $s^{(m)}(e,v)$ concave in e,v such that for any strongly r_o regular sequence $U_j \subset R^\nu$, with $|U_j|/j \longrightarrow v$, we have

$$\underset{j \longrightarrow \infty}{\text{limit}} \; s^{(m)}(j,U_j,e) = s^{(m)}(e,v)$$

(ii) There exists a function $f^{(c)}(\beta,v)$ convex in β, concave in v such that for any strongly r_o-regular sequence $U_j \subset R^{\nu}$ with $|U_j|/j \longrightarrow v$ we have

$$\lim_{j \longrightarrow \infty} f^{(c)}(j,U_j,\beta) = f^{(c)}(\beta,v)$$

(iii) There exists a function $p^{(g)}(\beta,\mu)$, convex in β,μ such that for any strongly r_o-regular sequence $U_j \subset R^{\nu}$, we have

$$\lim_{j \longrightarrow \infty} p^{(g)}(U_j,\beta,\mu) = p^{(g)}(\beta,\mu).$$

As usual, the proof proceeds by way of a number of lemmas. In all, the hypotheses of 9.27 will be assumed.

9.28 Lemma. Let C_j be a cube with side $j\,L - r_o/2$ $(L \geq r_o, j=1,2,\cdots)$. Then there are functions

$$s^{(m)}(e,v)\ ,\ f^{(c)}(\beta,v),\ p^{(g)}(\beta,\mu)$$

such that

(i) if $n(j)$ is an integer such that $|C_j|/n(j) \longrightarrow v$,

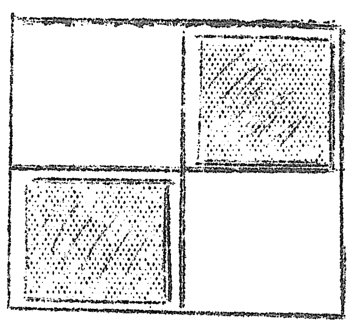
$$\lim_{j \longrightarrow \infty} s^{(m)}(n(j),C_j,e) = s^{(m)}(e,v)$$

(ii) for $n(j)$ as in (i),

$$\lim_{j \longrightarrow \infty} f^{(c)}(n(j),C_j,\beta) = f^{(c)}(\beta,v)$$

(iii) $\displaystyle\lim_{j \longrightarrow \infty} p^{(g)}(C_j,\beta,\mu) = p^{(g)}(\beta,\mu).$

Proof. By 9.23, $-s^{(m)}(n(j),C_j,e) \cdot n(j)$ is subadditive in j.



j_1 $\qquad$ j_2

By 9.22, $-s^{(m)}(n(j),C_j,e)$ is bounded below. Hence it converges. The other parts are similar. □

The next observation is important and seems to have been overlooked. (i.e. : it is often assumed, without justification that $s^{(m)}$ depends only on the density $|U|/n = \mathfrak{p}$).

9.29 Lemma. The functions defined by 9.28 are independent of L, n(j).

Proof. Each part is similar, so we prove it only for $p^{(g)}$. Suppose C_j and D_j are two such sequences of cubes with $C_j \subseteq D_j$. Given a constant K there are subsequences such that

$$(|D_j| - |C_j|\, K)/|D_j| \longrightarrow 0.$$

It is sufficient to check subsequences, for by 9.28 $p^{(g)}(C_j,\beta,\mu)$ and $p^{(g)}(D_j,\beta,\mu)$ both converge. Suppose $p^{(g)}(C_j,\beta,\mu) \longrightarrow a$ and write

$$|p^{(g)}(D_j,\beta,\mu) - a\,| \leq \frac{1}{|D_j|}[\log\mathcal{Q}(D_j,\beta,\mu) - \log\mathcal{Q}(C_j,\beta,\mu)]$$
$$+ \left|\frac{1}{|C_j|}\log\mathcal{Q}(C_j,\beta,\mu) - a\,\right|$$

The second term $\longrightarrow 0$ and by 9.23, 9.22 there are constants K_1,K_2 so that $\log\mathcal{Q}(D_j,\beta,\mu) \leq K_1|D_j|$ and $\log\mathcal{Q}(C_j,\beta,\mu) \geq K_2|C_j|$. Assume j large so we may take $K_1, K_2 > 0$. Hence we have the result.□

9.30 Lemma. The functions defined by 9.28 have the convexity and concavity properties stated in 9.27.

Proof. Three of these have already been proven (c.f. 9.18, 9.19). The rest are similar, so we prove that $s^{(m)}$ is concave in e. In 9.23, Suppose

$$e = \alpha\, e_1 + (1-\alpha)\, e_2 \quad \text{and choose} \quad e' = \frac{m\,\alpha}{n+m}\, e$$

so that

$$s^{(m)}(n_1+m, U_{n+m}, e) \geq \frac{n}{n+m} s^{(m)}(n, U_n, \alpha\, e_1) + \frac{m}{n+m} s^{(m)}(m, U_m, (1-\alpha\,)\, e_2)$$

Letting $n_1 m \longrightarrow \infty$ gives the result. □

Proof of 9.27. The proofs of (i), (ii) and (iii) are similar, so we prove only (i). Since $U_j/j \longrightarrow v$ we may work with $\frac{1}{|U_j|} \log \Sigma\ (j, U_j, e) = \tilde{s}^{(m)}(U_j, e)$ rather than $\frac{1}{j} \log \Sigma\ (j, U_j, e)$. The proof now proceeds in two steps which, together imply the theorem.

Step 1. $\liminf\ \tilde{s}^{(m)}(j, U_j, e) \geq \tilde{s}^{(m)}(e, v)$.

To see this, given L, choose T_j as in 9.26 and construct T_j' with edges $L - r_o/2$. Hence, by 9.22, if C_L denotes a component cube,

$$\tilde{s}^{(m)}(j, U_j, e) \geq \frac{|T_j'|}{|U_j|} \tilde{s}^{(m)}(n(L), C_L, e)$$

where $j = \Sigma\ n(L)$. Letting $j \longrightarrow \infty$ we see

$$\liminf\ \tilde{s}^{(m)}(j, U_j, e) \geq s^{(m)}(n(L),\ C_L, e)$$

now let $L \longrightarrow \infty$ and use 9.29.

Step 2. $\limsup\ \tilde{s}^{(m)}(j, U_j, e) \leq s^{(m)}(e, v)$

Here, given L, construct W_j as in 9.26 and the corresponding W_j'. Let C_j be a cube so $C_j \supset W_j \supset U_j$ and filling out C_j / W_j with cubes Y_j we have, by 9.23

$$|C_j|\tilde{s}^{(m)}(n(j),\ C_j, e) \geq |U_j|\tilde{s}^{(m)}(j, U_j, e) + |Y_j'|\ \tilde{s}^{(m)}(n(j)-j,\ CL, e)$$

Hence

$$\tilde{s}^{(m)}(j, Uj, e) \leq \frac{|C_j|}{|U_j|}\ \tilde{s}^{(m)}(n(j), C_j, e) - \frac{|Y_j'|}{|U_j|} \tilde{s}^{(m)}(n(j)-j, C_L, e)$$

Letting $L, j \longrightarrow \infty$ gives the result, since

$$\frac{|C_j|}{|U_j|} - \frac{|Y'_j|}{|U_j|} \longrightarrow 1.$$

(we leave the ϵ's to the reader.) Thus the proof is complete. $\square$

As Fisher points out, this argument is quite delicate and it is easy to be fallacious. We would therefore appreciate any comments on this theorem. (9.27).

§10. The Thermodynamic Limit of the Correlation Functions.

The basic theorem of this section is that the correlation functions have a thermodynamic limit $(U_n \subset R^\nu \longrightarrow \infty)$ and are analytic for $|z| < \varepsilon$ where $\varepsilon > 0$ depends on the interaction potential V. References are Groeneveld [1], Ruelle [2, (Ch. III), 3,4,5] and Penrose [1]. Our version is a slight simplification and generalization of part of Ruelle [3].

We begin with the definition of the correlation functions;

10.1 Definition. Consider a stable system with Hamiltonian

$$H^n : M^n \subset R^{n\nu} \times R^{n\nu} \longrightarrow R$$

$$H^n(q,p) = \sum_{i=1}^{n} p_i^2/2m + V_n(q).$$

As usual, for $|U| < \infty$, $\beta > 0$, $z \in C$,

$$\mathcal{Z}(U,z,\beta) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \int_{U^n \times R^{n\nu}} \exp(-\beta H^n) d\mu_n$$

$$= \sum_{n=0}^{\infty} \left(\frac{z}{\lambda^\nu}\right)^n \frac{1}{n!} \int_{U^n} \exp(-\beta V_n) dq_n$$

where $U^n = U \times U \times \dots \times U \cap M_q^n$, $M^n = M_q^n \times R^{n\nu}$.

The correlation functions are defined by

$$\rho_n^U(z,\beta;q_1,\dots,q_n) = \frac{C_{U^n}}{\mathcal{Z}(U,z,\beta)} \sum_{s=0}^{\infty} \{\frac{z^{n+s}}{s!} \times$$

$$\int_{U^s \times R^{(n+s)\nu}} \exp(-\beta H_{n+s}) dq_{n+1} \dots dq_{n+s}\, dp_1 \dots dp_{n+s}\}$$

$$= \frac{C_{U^n}}{\mathcal{Q}(U,z,\beta)} \sum_{s=0}^{\infty} \left(\frac{z}{\lambda^{\nu}}\right)^{n+s} \frac{1}{s!} \int_{U^s} \exp(-\beta V_{n+s}) dq_{n+1} \cdots dq_{n+s}$$

for $q_i \in R^{\nu}$, $p_i \in R^{\nu}$, where C_{U^n} is the characteristic function of U^n. (We allow ρ to assume the value $+\infty$.)

If, in a particular discussion, β (or z) is fixed throughout, it will be omitted from the symbols.

The two main theorems are as follows:

10.2 Theorem. In 10.1, suppose $V_n(q) = \sum_{i<j} V(q^i - q^j)$ where $V(q) = V(-q)$ and

(i) the system is stable,

and (ii) either the system is strongly tempered (§9) or the condition

$$(R) : C = \int_{R^{\nu}} [\exp(-\beta V(q)) - 1] dx < \infty$$

holds. Then there is an $\varepsilon > 0$ such that for $U_i \subset R^{\nu}$ with $\{\text{diameter } (U_i)\} \longrightarrow \infty$, we have

$$\rho_n^{U_i} \longrightarrow \rho_n \qquad n = 1,2,\ldots$$

uniformly in q on bounded sets, and z for $|z| < \varepsilon$. Moreover $\rho_n^{U_i}$ and ρ_n are analytic in z for $|z| < \varepsilon$. (The choice of ε depends on V).

10.3 Theorem. Under the hypotheses of 10.2,

$$\frac{1}{|U_i|} z \frac{\partial}{\partial z} \log \mathcal{Q}(U_i,z,\beta) \quad \text{and} \quad \frac{1}{|U_i|} \log \mathcal{Q}(U_i,z,\beta)$$

are analytic in some circle $|z| < \varepsilon$ and converge to an analytic function as $i \longrightarrow \infty$.

The rest of this section is devoted to proofs of these theorems. Using Ruelle's ideas, the proofs are remarkably easy, and elegant.

First we make the following observation:

10.4 <u>Lemma</u>. In 10.1, if $|U| < \infty$, $\mathcal{Q}(U,z,\beta)$ is analytic in z, and if $\mathcal{Q}(U,z,\beta)$ is not zero in some region (one always exists about the origin as $\mathcal{Q}(U,0,\beta) = 1$, and $\mathcal{Q}$ is continuous), then $\rho_n^U(z,q_1,\ldots,q_n)$ is analytic in z. Also, we have

$$\frac{1}{|U|}\int_U \rho_1^U(z,q)dq = z\frac{\partial}{\partial z}\log\mathcal{Q}(U,z)$$

$$= 1/v\ (U,z)$$

in this region.

<u>Proof</u>. The first part is immediate from stability, as in 8.2, and the second follows by integrating the expression for ρ_1 term by term, admissible as the series converges uniformly in q, for each z, by stability (c.f. Weierstrass M-test). $\square$

The idea of Ruelle was to cast the expressions for ρ_n into operator language in an appropriate Banach space, and apply standard powerful tools to extract the result. Our method differs from his in that we use the fixed point theorem rather than inverting an operator.

In pursuit of this result we derive identities for the correlation functions going back to Kirkwood-Salsburg [1] and Mayer-Montroll [1]. (See also Hill [1, p. 251-3].) (We only shall require the first identity below, but we give both for completeness).

10.5 <u>Theorem.</u> In 10.1, suppose the system is stable and $V_n(q) = \sum_{i<j} V(q_i - q_j)$; $q = (q_1, \ldots, q_n) \in R^{n\nu}$, $V(q_i) = V(-q_i)$. Let $z \in C$, $U \subset R^\nu$ and $\beta > 0$ be fixed. Define

$$f : R^\nu \longrightarrow R \;;\; f(x) = \exp[-\beta V(x)] - 1$$

$$K_n(x, x_1, \ldots, x_n) = \prod_{j=1}^{n} f(x - x_j)$$

and

$$K_{mn}(x_1, \ldots, x_m, y_1, \ldots, y_n) = \prod_{j=1}^{n} \prod_{k=1}^{m} f(y_j - x_k) \,.$$

Then we have

(i) (Kirkwood-Salsburg)

$$\rho_1(x_1) = C_U(x_1)\left(\frac{z}{\lambda^\nu}\right)\{1 + \sum_{s=1}^{\infty} \frac{1}{s!} \int_{U^s} K_s(x_1, x_2, \ldots, x_{s+1}) \rho_s(x_2, \ldots, x_{s+1}) dx_2 \ldots dx_{s+1}\}$$

and, for $n \geq 2$,

$$\rho_n(x_1, \ldots, x_n) = \left(\frac{z}{\lambda^\nu}\right) \exp\,[-\beta V_n^{(1)}(x_1, \ldots, x_n)] \times C_{U^n}(x_1, \ldots, x_n)$$

$$\{\rho_{n-1}(x_2, \ldots, x_n) + \sum_{t=1}^{\infty} \frac{1}{t!} \int_{U^t} K_t(x_1, x_{n+1}, \ldots, x_{n+t}) \rho_{n+t-1}(x_2, \ldots, x_{n+t}) dx_{n+1} \ldots dx_{n+t}\}$$

where $$V_n^{(1)}(x_1, \ldots, x_n) = \sum_{j=2}^{n} V(x_1 - x_j) \,.$$

(ii) (Mayer-Montroll) for all $n, m = 1, 2, 3, \ldots$ we have

$$\rho_m(x_1, \ldots, x_m) = \left(\frac{z}{\lambda^\nu}\right) \exp[-\beta V_m(x_1, \ldots, x_m)] \times$$

$$\{1 + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{U^n} K_{mn}(x_1, \ldots, x_m, y_1, \ldots, y_n)\; \rho_n(y_1, \ldots, y_n) dy_1 \ldots dy_n\}$$

<u>Proof</u>. (A simple exercise in combinatorics). Notice that $(n \geq 2)$

$$V_{n+s}(x_1,\ldots,x_{n+s}) = \sum_{j=2}^{n} V(x_1-x_j) + \sum_{j=n+1}^{n+s} V(x_1-x_j) + V_{n+s-1}(x_2,\ldots,x_{n+s})$$

so that

$$\exp[-\beta V_{n+s}(x_1,\ldots,x_{n+s})]$$

$$= \exp[-\beta V_n^{(1)}(x_1,\ldots,x_n)]\ \exp[-\beta V_{n+s-1}(x_2,\ldots,x_{n+s})] \prod_{j=n+1}^{n+s} (1+f(x_1-x_j))$$

Inserting this in the definition of ρ_n and expanding the last factor, noting symmetry gives

$$\rho_n(x_1,\ldots,x_n) = \frac{C_U^n}{\mathcal{Q}} \exp[-\beta V_n^{(1)}(x_1,\ldots,x_n)] \times$$

$$\sum_{s=0}^{\infty} \left(\frac{z}{\lambda^v}\right)^{n+s} \frac{1}{s!} \int_{U^s} \exp[-\beta V_{n+s-1}(x_2,\ldots,x_{n+s})] \times$$

$$\left\{\sum_{t=0}^{s} \frac{s!}{(s-t)!t!}\ f(x_1-x_{n+1})\ldots f(x_1-x_{n+t})\right\} dx_{n+1}\ldots dx_{n+s}$$

(the $t=0$ term being 1).

Since the terms are positive and we have convergence, we may rearrange the sums. Briefly,

$$\sum_{s=0}^{\infty} \sum_{t=0}^{\infty} = \sum_{t=0}^{\infty} \sum_{s=t}^{\infty}$$

Introducing $k=s-t$ then gives

$$\rho_n(x_1,\dots,x_n) = \frac{C_U^n}{\mathcal{Q}} \exp[-\beta V_n^{(1)}(x_1,\dots,x_n)] \times$$

$$\sum_{t=0}^{\infty} \sum_{k=0}^{\infty} \left(\frac{z}{\lambda^\nu}\right)^{n+k+t} \int_{U^{t+k}} \{\exp[-\beta V_{n+k+t-1}(x_2,\dots,x_{n+k+t})]$$

$$\frac{1}{k!t!} K_t(x_1,x_{n+1},\dots,x_{n+t})\} dx_{n+1}\cdots dx_{n+k+t} \ .$$

Removing a factor z/λ^ν, the remaining sum over k, with $\mathcal{Q}$ included is exactly

$$\rho_{n+t-1}(x_2,\dots,x_{n+t}).$$

The case $n=1$ is proven in a similar way.

For (ii) we proceed as above, using instead

$$V_{m+n}(x_1,\dots,x_m,x_{m+1},\dots,x_{m+n}) = V_m(x_1,\dots,x_m) + V_n(x_{m+1},\dots,x_{m+n})$$
$$+ \sum_{j=1}^{m} \sum_{k=m+1}^{m+n} V(x_j-x_k).$$

We leave the rest to the reader. □

Notice that $\mathcal{Q}$ has disappeared in these formulae. In (i), $n=1$ the first term would have been $\mathcal{Q}$ if we had not included it in the definition of ρ_n. This feature is essential to make the procedure work.

We will interpret the $\{\rho_n(z,x_1,\dots,x_n)\}$ as elements of a Banach space (U,β fixed). We now set up the machinery for this purpose.

10.6 Definition. Let (X,Σ,μ) be a measure space and $L^\infty(X,\mu)$ denote the (equivalence classes of) measureable functions $f : X \longrightarrow R$ such that $\text{ess sup } f \equiv \inf\{M \in R : |f(x)| \leq M \text{ almost everywhere}\} < \infty$.

The following is standard, but we give it anyway.

10.7 <u>Lemma</u>. $L^\infty(X,\mu)$ with the norm $\|f\|_\infty = \text{ess sup} f$ is a Banach space (B-space)

<u>Proof</u>. The properties $\|f\|_\infty = 0 \Longrightarrow f = 0$ a.e. and $\|af\|_\infty = a\|f\|_\infty$ are clear. For the triangle inequality, if $|f(x)| \leq M$ a.e. and $|g(x)| \leq N$ a.e. then $|f+g| \leq M + N$ a.e. Taking the infimum over M and N gives $\|f+g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$.

For completeness (see 2.3 for definitions), let f_n be a Cauchy sequence. Then $f_n(x)$ converges for almost every x, defining a measureable function f. We must show that $\|f_n - f\|_\infty \longrightarrow 0$ (this implies $f \in L^\infty(X,\mu)$ and $f_n \longrightarrow f$). Given $\varepsilon > 0$ choose N so $n,m \geq N$ implies $\|f_n - f_m\|_\infty < \varepsilon/2$. For $k \geq N$ and any $\ell \geq N$ we have

$$|f_k(x) - f(x)| \leq |f_k(x) - f_\ell(x)| + |f_\ell(x) - f(x)|$$

$$< \varepsilon/2 + |f_\ell(x) - f(x)|$$

But for a.e. x, choose $\ell \geq L$ (which depends on x) so $|f_\ell(x) - f(x)| \leq \varepsilon/2$. Thus for any $x \in X$, $|f_k(x) - f(x)| < \varepsilon$ for $k \geq N$. $\square$

We shall also require the following well known fact from complex analysis.

10.8 <u>Lemma</u>. The bounded analytic functions $f : D \subset \mathbb{C} \longrightarrow \mathbb{C}$ (D open) form a Banach space with

$$\|f\| = \sup\{|f(z)| : z \in D\}.$$

The next lemma is another routine verification much like 10.7, and so is omitted.

10.9 Lemma. Suppose $E_1, E_2, \ldots$ are Banach spaces with norms $\| \|^{(1)}, \| \|^{(2)}, \ldots$ and $\sigma = (\sigma_1, \sigma_2, \ldots)$ is a sequence of positive real numbers. Let

$$E = \bigoplus_{i=1}^{\infty} E_i$$

(sequences of elements $e = (e_1, e_2, \ldots)$) and

$$\hat{E} = \{e \in E : \|e\|_\sigma = \sup \{\sigma_n \|e_n\|^{(n)}\} < \infty\} .$$

Then $\hat{E}$ is a Banach space with norm $\| \|_\sigma$.

One of our basic tools below will be the following

10.10 Lemma (Contraction Lemma). Let M be a complete metric space (1.8) and $f : M \longrightarrow M$ a mapping. Suppose there is a constant κ so $0 \leq \kappa < 1$, and

$$d(f(x), f(y)) \leq \kappa \, d(x,y)$$

for all $x, y \in M$. Then f has a unique fixed point (a point such that $f(x_0) = x_0$). The map f is continuous.

Proof. Let $x \in M$ and consider the sequence $\{f^{(n)}(x)\}$ of the iterates of f applied to x. Then $d(f^{(n)}(x), f^{(n+p)}(x))$

$$\leq d(f^{(n)}(x), f^{(n+1)}(x)) + \ldots + d(f^{(n+p-1)}(x), f^{(n+p)}(x))$$

$$\leq (\kappa^n + \ldots + \kappa^{n+p-1}) \, d(x, f(x))$$

by the triangle inequality. Since $1+\kappa+\kappa^2+\ldots$ is convergent, the sequence $f^{(n)}(x)$ is Cauchy. Since M is a complete metric space, the Cauchy sequence $f^{(n)}(x)$ has a limit point y. This limit point y is a fixed point of f since

$$\|f(y) - y\| \leq \|f(y) - f^{(n+1)}(y)\| + \|f^{(n+1)}(y) - y\|$$

$$\leq \kappa \|f^{(n)}(y) - y\| + \|f^{(n+1)}(y) - y\|$$

and the right hand side approaches zero as $n \longrightarrow \infty$. This fixed point y is unique, for $f(x) = x$ and $f(y) = y$, then

$$d(x,y) = d(f^{(n)}(x), f^{(n)}(y)) \leq \kappa^n d(x,y) .$$

Hence $d(x,y) = 0$, or $x = y$. Continuity of f is obvious. □

We now return to the correlation functions.

10.11 <u>Definition</u>. In 10.5, let E_n denote the Banach space of functions $\varphi_n(z,x_1,\ldots,x_n)$ analytic and bounded in z for $|z| < \varepsilon$ and essentially bounded in $x_1,\ldots,x_n \in R^\nu$. Given $\xi > 0$ let $\sigma = (\xi,\xi^{-1},\xi^{-2},\ldots)$ and

$$E = \bigoplus_{i=1}^{\infty} E_i$$

(with the usual vector space structure) and let

$$E_{\xi,\varepsilon} = \hat{E}_\sigma, \quad \text{norm} \quad \| \ \|_{\xi,\varepsilon}$$

For
$$\varphi \in E_{\xi,\varepsilon} , \ \varphi = (\varphi_1,\varphi_2,\ldots),$$

define (β fixed)

$$(K_U\varphi)_n(z,x_1,\ldots,x_n) , \ n = 1,2,\ldots$$

by

$$(K_U\varphi)_1(z,x_1) = C_U(x_1)\,\frac{z}{\lambda^\nu}\,\{1 + \sum_{s=1}^{\infty}\frac{1}{s!}\times$$

$$\int_{U^s} K_s(x_1,\ldots,x_{s+1})\,\varphi_s(z,x_2,\ldots,x_{s+1})dx_2\ldots dx_{s+1}$$

and
$$(K_U\varphi)_n(z,x_1,\ldots,x_n) = C_{U^n}(x_1,\ldots,x_n)\,\frac{z}{\lambda^\nu}\times$$

$$\exp[-\beta V_n^{(1)}(x_1,\ldots,x_n)]\,\{\varphi_{n-1}(x_2,\ldots,x_n) + \sum_{t=1}^{\infty}\frac{1}{t!}\times$$

$$\int_{U^t} K_t(x_1,x_{n+1},\ldots,x_{n+t})\,\varphi_{n+t-1}(x_2,\ldots,x_{n+t})dx_{n+1}\ldots dx_{n+t}\}$$

where $V_n^{(1)}$, K_t are defined as in 10.5, and we regard $1 \in E_1$.

We next wish to show that for ε, ξ appropriately chosen, $K_U : E_{\xi,\varepsilon} \longrightarrow E_{\xi,\varepsilon}$ and is a contraction (satisfies 10.10). Warning : K_U is not a linear map. (Ruelle's corresponding operator is linear; Ruelle [3]).

To do this we require condition (R) stated in 10.2. The hypothesis of strong tempering is then eliminated by the following.

10.12 Lemma. In 10.2, if the system is stable and strongly tempered (§9) then for any $\beta > 0$,

$$C = \int_{R^\nu} (\exp[-\beta V(x)]-1)dx < \infty\ .$$

Although the proof is straightforward it is rather awkward. For the details, see Ruelle [2, p. 90]. A trivial proof of this lemma probably exists.

We shall therefore be content to prove 10.2 under stability, and condition R.

10.13 Lemma. In 10.11, suppose V is stable and satisfies (R) of 10.2. Then if $\xi = 1/C$ and $\varepsilon < \frac{\lambda^\nu}{C}\exp(-2\beta B-1)$, then $K_U : E_{\xi,\epsilon} \longrightarrow E_{\xi,\varepsilon}$ and is a contraction for any $U \subset R^\nu$. More generally, for any $\xi > 0$, we may choose $\varepsilon < \xi\, e^{-2\beta B} e^{-\xi C}$.

Proof. From stability and permutation symmetry, we see that $V_m^{(1)}(x_1,\dots,x_m) \geq -2B$; namely, $\sum_{i=1}^{n} \sum_{j\neq i} V(x_i-x_j) \geq -2nB$ by definition of stability, and so there is an i, so $\sum_{j\neq i} V(x_i-x_j) \geq -2B$.

Next, observe that

$$\xi^{-m}\|(K_U\varphi)_m\|_\infty \leq \frac{\varepsilon}{\lambda^\nu}\cdot \exp(-\beta B)\cdot \xi^{-m}$$

$$\{\|\varphi_{m-1}\|_\infty + \sum_{n=1}^{\infty} \frac{1}{n!}\|\varphi_{m+n-1}\|_\infty \cdot C^n\}$$

$$\leq \frac{\varepsilon}{\lambda^\nu}\exp(2\beta B)\ \{\xi^{-1}\|\varphi\|_{\xi,\varepsilon} + \xi^{-1}\|\varphi\|_{\xi,\varepsilon} \times$$

$$\sum_{n=1}^{\infty}\frac{1}{n!}C^n\cdot\xi^n\}$$

using the fact that $\xi^{-m}\|\varphi_m\|_\infty \leq \|\varphi\|_{\xi,\varepsilon}$. Hence

$$\xi^{-m}\|(K_U\varphi)_m\|_\infty \leq \frac{\varepsilon}{\lambda^\nu}\exp(2\beta B)\ \xi^{-1}\cdot\|\varphi\|_{\xi,\varepsilon}\cdot\exp(C\xi)$$

$$= \frac{\varepsilon}{\lambda^\nu}\exp(2\beta B+1)\ \xi^{-1}\|\varphi\|_{\xi,\varepsilon}$$

$$< \|\varphi\|_{\xi,\varepsilon}.$$

Hence $K_U\varphi \in E_{\xi,\varepsilon}$. Moreover, K_U is a contraction as $d(K_U\varphi, K_U\psi) = \|K_U\varphi - K_U\psi\|_{\xi,\varepsilon}$ and the first (constant) terms cancel, so the above inequalities apply. □

10.14 <u>Lemma</u>. In the above, the fixed point of K_U is the vector where components are the correlation functions $\{\rho_n^U(z,x_1,\ldots,x_n)\}$. In particular, ρ_n is analytic for $|z| < \varepsilon$.

<u>Proof</u>. The ρ_n^U satisfy the Kirkwood-Salsburg equations, and so are fixed points of K_U. The only question is whether or not $\rho_n^U \in E_{\xi,\varepsilon}$. However, for ε sufficiently small, ρ_n^U is analytic so coincides with the fixed point of K_U on a (smaller) neighborhood of $z = 0$. Since ρ_n^U has at most a finite number of singularities in $|z| < \varepsilon$, we can repeat on all open discs where there are no singularities, to conclude that the singularities are in fact removeable. Hence the lemma. □

The final basic property of the operators K_U we need is as follows.

10.15 <u>Lemma</u>. Consider the operators K_{U_i} on functions on a bounded set $S \subset R^\nu$. Then there is a $\xi > 0$ and corresponding $\varepsilon > 0$ such that

$$K_{U_i} \longrightarrow K$$

uniformly, on the fixed points.

<u>Proof</u>. First we claim that all the fixed points $\rho_n^{U_i}$ are uniformly bounded. In fact, from before, we see that ξ,ε can be chosen so that $\|K_{U_i}\varphi\|_\xi \le \max(1,\|\varphi\|_\xi)$ and so if we start with a $\|\varphi\|_\xi \le 1$, and note $\rho^{U_i} = \lim_{n\to\infty} K_{U_i}^n \varphi$ then clearly $\|\rho^{U_i}\|_\xi \le 1$. (independent of U_i). Secondly, we claim that $K_{U^i} \longrightarrow K$ uniformly on any bounded set, and the fixed points in particular. In fact, from the definition 10.11, choose U_i so large that $U_i \supset S$, and we can

set $C_{U_i} = 1$. (This is why bounded sets are required.) Then since K_t is integrable, given $\delta > 0$, choose i so large that

$$\int_{R^\nu \setminus U_i} K_t(x_1, x_{m+1}, \ldots, x_{m+t}) dx_{m+1} \ldots dx_{m+t} \leq (\delta)^t$$

Then

$$\xi^{-m} \| (K_{U_i} \varphi - K\varphi)_m \|_\infty \leq \frac{\varepsilon}{\lambda^\nu} \exp(-2\beta B) \{ \sum_{t=1}^{\infty} \frac{1}{t!} \times$$

$$(\int_{(R^\nu \setminus U)^t} K_t(x_1, x_{m+1}, \ldots, x_{m+t}) dx_{m+1} \ldots dx_{m+t}) \ \xi^{t-1} \|\varphi\|_\xi \}$$

$$\leq \frac{\varepsilon}{\lambda^\nu} \exp(2\beta B) \ \{ \xi^{-1} \{\exp(\delta\xi) - 1\} \ \|\varphi\|_\xi \}$$

This proves the result. $\square$

One more lemma and we are at our goal!

10.16 Lemma. Let M be a complete metric space and $f_i, f : M \longrightarrow M$ contractions on M with the same constant κ. Let x_i be the fixed point of f_i and suppose $f_i \longrightarrow f$ uniformly on these fixed points. Then $\{x_i\}$ converges to the fixed point of f.

Proof. First, note that x_i is a cauchy sequence. In fact, given $\varepsilon > 0$, choose N so $m, n \geq N$ implies $d(f_m(x_i), f_n(x_i)) < \varepsilon(1-\kappa)$ for all i. Then

$$d(x_n, x_m) \leq d(f_n(x_n), f_m(x_n))$$

$$+ d(f_m(x_n), f_m(x_m))$$

$$\leq \varepsilon(1-\kappa) + \kappa d(x_n, x_m) .$$

Hence $d(x_n,x_m) \leq \varepsilon$.

Now suppose that $x_i \longrightarrow x$, so that $f_n(x_n) \longrightarrow x$. Given $\varepsilon > 0$, choose N so $n \geq N$ implies $d(f_n(x_n),f(x_n)) < \varepsilon/2$ and $d(f_n(x_n),x) < \varepsilon/2$. Then $d(f(x_n),x) < \varepsilon$. Therefore $f(x_n) \longrightarrow x$ and $f(x_n) \longrightarrow f(x)$ since f is continuous. Thus $x = f(x)$, the required result. □

<u>Proof of 10.2</u>. By 10.15 the contractions $C_s K_{U_i}$ (with fixed points $C_s\varphi^{U_i}$), are uniformly convergent on the fixed points. Hence, by 10.16, 10.13 $C_s\varphi^{U_i} \longrightarrow C_s\varphi$ (uniformly), which is the result. Note that the contraction constants are independent of U_i. Analyticity was given in 10.14. □

<u>Proof of 10.3</u>. Clearly the correlation functions in the limit are translation invariant. In particular $\rho_1(x_1)$ is constant, say $1/v$. Then by 10.4,

$$\left| \frac{z}{|U_i|} \frac{\partial}{\partial z} \log \mathcal{Q} - \frac{1}{v} \right| = \frac{1}{|U_i|} \int_{U_i} [\rho_1^{U_i}(x_1) - C_{U_i}\rho_1(x_1)]dx_1$$

From the proof of 10.15, given $\varepsilon > 0$, there is an N so $i \geq N$ implies $|\rho_1^{U_i} - \rho_1| < \varepsilon$ on U_i. Hence we have

$$\left| \frac{z}{|U_i|} \frac{\partial}{\partial z} \log \mathcal{Q} - \frac{1}{v} \right| < \varepsilon,$$

as required. Since the functions are analytic and uniformly convergent in z, the second part follows as well. □

In this development we have a lower bound (ε) on the region of analyticity, (but probably a crude one). Penrose [1] gives, on the other hand an upper bound.

The analyticity established here justifies the classical virial expansions for sufficiently low density (see §7).

The correlation functions where existence has been established above have an important general feature: the so-called <u>cluster decomposition property</u>. This says that if $x_1 \ldots x_n$ are partitioned into clusters $x_{\alpha_1} \ldots x_{\alpha_{n_1}} \ldots x_{\gamma_1} \ldots x_{\gamma_{n_k}}$ with $\sum_{\gamma=1}^{k} n_j = n$ and the clusters are made to separate from each other, then

$$\lim \rho(x_1 \ldots x_n) = \lim \rho(x_{\alpha_1} \ldots x_{\alpha_{n_1}}) \ldots \rho(x_{\gamma_1} \ldots x_{\gamma_{n_k}}) .$$

This behavior is appropriate to that of a pure phase. As was pointed out in Mayer and Montroll [1], when two distinct phases are present, this cluster decomposition property does not hold.

In fact, the existence theorems of this chapter apply only to the gas phase and it is a major unsolved problem of the subject to extend the theory beyond the first phase transition. The farthest the formalism which uses correlation functions has been taken is illustrated in Ruelle [5]. The great difficulties encountered have led to the introduction of new formalism but the main problems remain open.

Bibliography for Chapter II.

R. Abraham and J. Marsden [1] Foundations of Mechanics. Benjamin, New York (1967).

L. Ahlfors [1] Complex Analysis. McGraw Hill, N.Y. (1966).

T. Apostol [1] Mathematical Analysis. Addison Wesley, Reading Mass. (1957).

J. Blatt and V. Weisskopf [1] Theoretical Nuclear Physics. J. Wiley, N.Y. (1952).

S. Bochner [1] Lectures on Fourier Integrals, Princeton Univ. Press. (1959).

J. Dorfman and E. Cohen [1] Difficulties in the Kinetic Theory of Dense Gases. Journ. Math. Physics 8 (1967) 282-297.

F. J. Dyson and A. Lenard, Stability of Matter I. Jour. Math. Phys. 8 (1967) 423-433.

M. Fisher [1] The Free Energy of a Macroscopic System. Archive for Rat. Mech. and Analysis 17 (1964) 377-410.

M. Fisher and D. Ruelle [1] Stability of Many Particle Systems. Jour. Math. Physics 7 (1966) 260-270.

H. Frisch and J. Lebowitz [1] The Equilibrium Theory of Classical Fluids. Benjamin, New York (1964).

R. Giffiths [1] A proof that the Free Energy of a Spin System is Extensive. Journ. Math. Physics 5 (1964).

J. Groeneveld [1] Two Theorems on Classical Many Particle Systems. Physics Letters 3 (1962) 32-3.

P. Halmos [1] Lectures on Ergodic Theory. Chelsea N. Y. (1956).
[2] Measure Theory. D. Van Nostrand N. Y. (1950).
[3] Introduction to Hilbert Space and Spectral Multiplicity, Chelsea N. Y. (1951).

T. Hill [1] Statistical Mechanics. McGraw Hill N. Y. (1956).

L. Van Hove [1] Quelque Propértés General de I'integral de Configuration d'un Systeme de Particles avec interaction. Physica 15 (1949) 951-961.

K. Huang [1] Statistical Mechanics. J. Wiley N. Y. (1963).

A. Khinchin [1] Mathematical Foundations of Statistical Mechanics. Dover, N. Y. (1949).

J. Kirkwood and Z. Salsburg [1], in. Discussions Faraday Soc. (chem). 15 (1953) 28.

E. Lieb [1] A New Method in the Theory of Imperfect Gases and Liquids, Journ. Math. Physics. 4 (1963) 671-8.

J. Van der Linden [1] Rigorous Results for some Thermodynamic Limit Problems. Physica 32 (1966) 642-668.

J. Mayer and M. Mayer [1] Statistical Mechanics. J. Wiley N.Y. (1940).

J. Mayer [2] Integral Equations between Distribution Functions of Molecules. Journ. Chem. Phys. 15 (1947) 187.

J. Mayer and E. Montroll [1] Molecular Distributions. Journ. Chem. Phys. 9 (1941) 2.

B. Mazur, On the Asymptotic Equivalence of various Ensembles for a real System. I.U.P.A.P. Conference on Stat. Mech. and Thermodynamics, Copenhagen (1966).

L. Page, Introduction to Theoretical Physics. Van Nostrand, Princeton, N.J. (1952).

O. Penrose, Fugacity Expansions for Fluids and Lattice Gases. Journ. Math. Phys. 4 (1963) 1312.

D. Ruelle [1] Classical Statistical Mechanics of a System of Particles, Helv. Phys. Acta. 36 (1963) 183-197.

[2] Rigorous Results in Statistical Mechanics, Lectures in Theoretical Physics (IV) 1963. Univ. of Colorado Press, Boulder (1964).

[3] Correlation Functions of Classical Gases, Ann. of Physics 25 (1963) 109-120.

[4] Cluster Property of Correlation Functions of Classical Gases. Rev. Mod. Physics. 36 (1964) 580-584.

[5] Correlation Functionals. Journ. Math. Physics 6 (1965) 201-216.

L. Schwartz [1] Theorie des Distributions. Tome 2, Hermann Paris (1959).

E. Titchmarsh [1] The Theory of Functions. Oxford Univ. Press. (1939)

C. Uhlenbeck and G. Ford [1] Lectures in Statistical Mechanics. A.M.S. publ. Providence R.I. (1963).

G. Whyburn [1] Topological Analysis. Princeton Univ. Press (1964).

C. N. Yang and T. D. Lee. Statistical Theory of Equations of State and Phase Transitions I. Theory of Condensation. Phys. Rev. 87 (1952) 404-409.

CHAPTER III

States in Classical and Quantum Statistical Mechanics.

Recent trends in statistical mechanics, particularly the approach using C^* algebras, rests heavily on a more refined notion of state. These ideas have been influenced greatly by quantum mechanics. The object of the treatment in these lectures, not quite achieved, was to show how the necessity for these new ideas arises naturally from elementary examples. Since the subject has many open problems, it is far from clear that a definitive formulation has been obtained but, surely, something is going on there.

In this chapter we summarize some of the elementary aspects of states, give a brief account of the current status of the subject, and give a resumé of the literature on the C^* algebra approach.

§11. States in Finite Dimensional Spaces.

Although the underlying spaces in Koopmanism and quantum mechanics are infinite dimensional, it is convenient to first see the formalism for finite dimensional spaces. The general case is similar; however non trivial pathologies of physical interest can and do arise.

11.1 Definition. If E,F are (finite dimensional) complex vector spaces, we let $L(E,F)$ denote the set of (continuous) linear maps $f : E \longrightarrow F$ with the structure of a complex vector space. We also write $L(E) = L(E,E)$. Let, in case E is provided with an inner product, $(\, , \,)$ $J(E) = \{A \in L(E) : (e_1, Ae_2) = (Ae_1, e_2)\}$, the self adjoint (symmetric) elements of $L(E)$, together with the structure of a <u>real</u> vector space. Elements of $J(E)$ are called <u>observables</u>. For $A \in L(E)$, A^*, the adjoint is the unique element of $L(E)$ satisfying $(e_1, Ae_2) = (A^* e_1, e_2)$ for all $e_1, e_2 \in E$. We also let $E^* = L(E, \mathbb{C})$, the dual.

11.2 <u>Theorem</u>. If $\dim E = n$, then $L(E)$ and $J(E)$ both have dimension n^2.

<u>Proof</u>. It is elementary linear algebra that $\dim L(E) = n^2$ as a complex vector space. Let $A_1, \ldots, A_m$ be a basis of $J(E)$. In $L(E)$ they are obviously linearly independent. Since, for each $A \in L(E)$, $A = \frac{1}{2}(A + A^*) + i\{\frac{1}{2i}(A - A^*)\}$, and $A + A^*$, $\frac{1}{i}(A - A^*) \in J(E)$, it is clear that they also span $L(E)$. $\square$

As a corollary, if $f \in L(J(E), R)$, then f has a unique extension to an element of $L(L(E), C)$, which we also denote f.

11.3 <u>Theorem</u>. $L(E)$ is a complex Hilbert space with the inner product $(A,B) = \operatorname{tr}(A^*B)$, and $J(E)$ is a real Hilbert space with $(A,B) = \operatorname{tr}(AB)$.

<u>Proof</u>. If $A \in L(E)$ we may identify, via a basis $A \in C^{n^2}$, and then

$$\operatorname{tr}(A^*B) = \sum_{i=1}^{n^2} \bar{a}_i b_i$$

which is an inner product. $\square$

Recall that $A \in L(E)$ is <u>positive</u>; $A \geq 0$ iff $(e, Ae) \geq 0$ for all $e \in E$. If $A \in J(E)$ then A is positive iff its eigenvalues are positive (≥ 0). It is a theorem of linear algebra (obvious for $A \in J(E)$) that $A \geq 0$ iff there is B so $A = B^*B$.

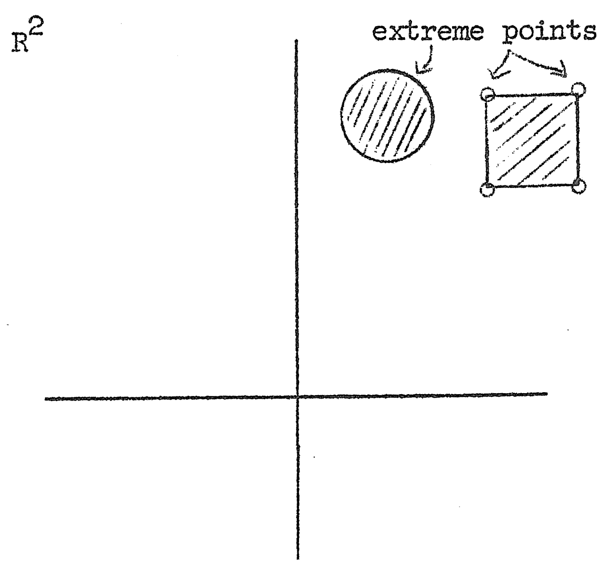
11.4 Definition. The states on E are defined by

$\mathcal{E} = \{f \in L(J(E),R) : f(1) = 1, \ A \geq 0 \Longrightarrow f(A) \geq 0\}$ where $1 \in J(E)$ is the identity map.

For $e \in E$, $\|e\| = 1$, let $f_e \in \mathcal{E}$, $f_e(A) = (e, Ae)$.

Clearly the extension $f \in L(L(E),C)$ is also positive as $A \geq 0$ implies $A \in J(E)$. As we shall see below, the states f_e are pure. (In statistical mechanics one is generally interested in impure states as well).

11.5 Definition. Let E be a linear space and $S \subset E$ a subset. Then S is called convex iff for all t, $0 \leq t \leq 1$, $e_1, e_2 \in S$ we have $te_1 + (1-t)e_2 \in S$. A point $e \in S$ is an extreme point in case $e_1 \neq e_2$, $e_1, e_2 \in S$, $e = te_1 + (1-t)e_2$, $0 \leq t \leq 1$ implies $t = 0$ or $t = 1$.



11.6 Theorem. The states $\mathcal{E}$ of E form a convex subset of $L(J(E),R)$. Also $\mathcal{E}$ is a (topologically) closed subset. The extreme points of $\mathcal{E}$ are called pure states.

Proof. (Any finite dimensional vector space is equipped uniquely with a topology determined by any norm). If $f_1, f_2 \in \mathcal{E}$ then $(tf_1 + (1-t)f_2)(1) = t + (1-t) = 1$ and if $A \geq 0$, $(tf_1 + (1-t)f_2)(A) \geq 0$ (R^+ is convex). Hence $\mathcal{E}$ is convex. To show $\mathcal{E}$ is closed we must show that if $f_n \longrightarrow f$, $f_n \in \mathcal{E}$ then $f \in \mathcal{E}$. But this is obvious. □

We now recall a basic theorem of Hilbert space theory which, in the finite dimensional case, is clear.

11.7 Theorem. (Riesz) Let $\mathcal{H}$ be a Hilbert space and $f \in \mathcal{H}^*$. Then there is an $e \in \mathcal{H}$ so $f(e') = (e,e')$ for all $e' \in \mathcal{H}$.

11.8 Definition. Let $f \in \mathcal{E}$. Then the unique $\rho_f \in J(E)$ so that $f(A) = \mathrm{tr}(\rho_f A)$ (by 11.7, 11.3), is called the density matrix of the state f.

Thus the association $f \longmapsto \rho_f$ is a bijection between $\mathcal{E}$ and $\mathcal{E}' = \{\rho \in J(E) : \mathrm{tr}\rho = 1, \rho \geq 0\}$ (taking $A(e_1) = (e,e_1)e$ we see $(e,\rho_f e) \geq 0$). Also, we see that the association is linear, when defined, so preserves convexity and extreme points.

11.9 Theorem. $f \in \mathcal{E}, f \neq 0$ is a pure state iff its density matrix ρ_f is a projection onto a one dimensional subspace, iff there is an $e \in E$, $\|e\| = 1$ so $f(A) = (e,Ae)$.

Proof. From elementary linear algebra (see, for example Hoffman and Kunze [1, p. 172]) we may write, uniquely, $\rho_f = \Sigma\, \lambda_i p_i$ where λ_i are the distinct eigenvalues and p_i are projections onto independent subspaces. If f is a pure state then ρ_f is an extreme point and so, since the eigenvalues add to one, the above sum can contain only one term and λ has multiplicity one. Conversely, if ρ_f is a projection and $\rho_f = t\,\rho_{f_1} + (1-t)\rho_{f_2}$ then by uniqueness of the spectral decomposition above, and the positivity of the eigenvalues, we see that if $t \neq 0,1$, $\rho_{f_1} = \rho_{f_2} = \rho_f$.

For the second equivalence, if $f(A) = (e,Ae)$ then ρ_f is the projection to e as $\mathrm{tr}(\rho_f A) = (e,Ae)$. Similarly we have the converse. □

$\rho \in L(E)$ is called <u>strictly positive</u> iff ρ is positive and $(x,\rho x) = 0$ implies $x = 0$. Note that $\rho \in J(E)$ is strictly positive iff its eigenvalues are.

11.10 <u>Theorem</u>. $\mathcal{E}'$ is a closed convex subset of $J(E)$, and of $J(E) \cap \{\rho \in J(E) : \mathrm{tr}\rho = 1\}$. With the relative topology of the later set, the interior of $\mathcal{E}'$ (topological interior; the largest open subset) consists of the strictly positive elements of $\mathcal{E}'$, and the boundary is $\partial\mathcal{E}' = \{\rho \in \mathcal{E}' : \det \rho = 0\}$. If $\dim E > 1$, the pure states are contained in $\partial\mathcal{E}'$.

<u>Proof</u>. That $\mathcal{E}'$ is closed is clear, as in 11.6. If $\rho \in \mathcal{E}'$ is strictly positive, which holds iff $\det \rho \neq 0$, then so is a neighborhood of ρ, since $\det$ is a continuous map. Hence $\rho \in$ interior $\mathcal{E}'$. If $\rho \in \mathcal{E}'$ is not strictly positive ($\det \rho = 0$) and $\varepsilon > 0$ there is a strictly positive ρ' so $\|\rho - \rho'\| < \varepsilon$ as is seen by diagonalizing for example. Hence the theorem. □

11.11 <u>Example</u>. If $\dim E = 2$, the pure states coincide with $\partial\mathcal{E}'$, but not if $\dim E > 2$.

Proof. If $\det \rho = 0$ and $\dim E = 2$ then ρ is a projection onto a one dimensional subspace (or $\rho = 0$) so is a pure state. The assertion for $\dim E > 2$ is obvious. $\square$

11.12 Definition. The entropy on $\mathcal{E}'$ is the map $s : \mathcal{E}' \longrightarrow R$; $s(\rho) = -\mathrm{tr}(\rho \log \rho)$.

Here $\log\rho$ is the unique map such that $\exp(\log\rho) = \rho$ and is defined on all complex matrices. As it is invariant under similarity, we can compute it easily for diagonalizable maps. In particular, if $\rho \in \mathcal{E}'$ then $\log\rho \in J(E)$ (ie : it remains real). See §5 for basic properties of $-x\log x$.

11.13 Theorem (von Neumann). The entropy s is strictly concave on $\mathcal{E}'$; that is for $\rho_1, \rho_2 \in \mathcal{E}'$,

$$s(\alpha\rho_1 + (1-\alpha)\rho_2) \geq \alpha s(\rho_1) + (1-\alpha)s(\rho_2)$$

for all α, $0 \leq \alpha \leq 1$ and equality holds iff $\rho_1 = \rho_2$ or $\alpha = 0$, or $\alpha = 1$.

The inequality is clear if ρ_1 and ρ_2 are simultaneously diagonalizable (i.e. commute). In general, however, it is a non-trivial inequality on the eigenvalues. (What makes it so delicate seems to be the fact that the eigenvalues of a sum of matrices is not related in any simple way to the separate eigenvalues). The proof may be found in von Neumann [1, Ch. 5, §3].

11.14 Corollary. Let $\mathcal{E}'_1 \subset \mathcal{E}'$ be a closed convex subset of $\mathcal{E}'$. Then there is a unique $\rho \in \mathcal{E}'_1$ which maximizes the entropy s. It is called the most chaotic state of ${}'_1$.

<u>Proof</u>. As was observed in §5, since $\operatorname{tr}\rho = 1$, $s(\rho\log\rho) = -\Sigma\, \lambda_i \log \lambda_i \leq \log n$; that is, s is bounded. Since s is continuous and $\mathcal{E}_1'$ is closed, s assumes its least upper bound. To see that it is unique, suppose it is assumed at ρ_1, ρ_2 and $\rho_1 \neq \rho_2$. But then if $\alpha \neq 0,1$ we have, by 11.13, $s(\alpha\rho_1 + (1-\alpha)\rho_2) > s(\rho_2)$. $\square$

A typical physical situation to which 11.14 applies is the following. Fix $A_1, \ldots, A_k \in J(E)$ and $a_1, \ldots, a_k \in R$ and let

$$\mathcal{E}_1' = \{\rho \in \quad ' : \operatorname{tr}(\rho A_i) = a_i \;;\; i = 1, \ldots, k\} \;.$$

Clearly $\mathcal{E}_1'$ is a closed convex subset. It corresponds to the situation in which we know the expectation values $a_1, \ldots, a_k$ of k observables $A_1, \ldots, A_k$. For further discussion along these lines, see Wichmann [1].

GLOSSARY OF SYMBOLS.

~~R reals~~
~~T^t flow 3~~

R	reals
Z	$\{\dots -2, -1, 0, 1, 2, \cdots\}$
T^t	flow 3
(X, Σ, μ)	measure space 10
A^c	complement 10
$A \setminus B$	set theoretic difference
$\overline{A}$	closure 14
∂A	topological boundary $= \overline{A} \cap \overline{A^c}$.
$X \times Y$	cartesian product
$L^p(X, \mu)$	19
$\Vert \quad \Vert$	norm 20
$\langle \ , \ \rangle$	inner product, 20
$\mathcal{H}$	Hilbert space, 20
U^t	induced flow, 20
$\hat{f}$	time average, 29
$\mathcal{A}$	partition, 36
$H(\mathcal{A})$	entropy, 36
$\vee$	common refinement, 37
h	Kolmogoroff-Sinai entropy, 38

$T_m M$	tangent space, 44
g	pseudo-Riemannian metric, 44
$H(q, p)$	Hamiltonien, 50
$V_{ij}(q)$	two body potential, 50
$\mathcal{F}(M)$	smooth functions, 50
d	exterior derivative, 51
P_X	momentum of X, 51
$\{f, g\}$	Poisson bracket, 51
T	temperature, 53
G_X	virial function of X, 54
p	pressure 56
κ	thermal conductivity, 60
Q_D	partition function for region D, 71
$\mathcal{Z}_D$	grand partition function, 73
z	fugacity, 73
P_U	pressure, 75
v_U	specific volume 75
$\lvert U \rvert$	volume of U, 75
$\Sigma(n, U, e)$	104
lc) $S^{(m)}(n, U, e)$	microcanonical entropy, 105
$\partial h / \partial U$	derivative with respect to scale, 107
$\partial h / \partial \lvert U \rvert$	derivative with respect to volume, 108

$\beta^{(m)}$	microcanonical temperature^{-1}, 108
$p^{(m)}$	microcanonical pressure, 109
$f^{(c)}$	canonical free energy, 109
$e^{(c)}$	canonical internal energy, 109
$p^{(c)}$	canonical pressure, 109
$p^{(g)}$	grand canonical pressure, 110
$v^{(g)}$	grand canonical specific volume, 110
$e^{(g)}$	grand canonical internal energy, 110
ρ_n^U	correlation functions for region U, 127
C	integral constant, 128
K_n	Kirkwood Salsburg function, 130
K_U	contraction mapping, 136
$E_{\xi,\epsilon}$	Banach space, 135
$L(E,F)$	continuous linear maps $f: E \to F$; ~~144~~ 144
$L(E)$	$L(E,E)$, 144
$J(E)$	symmetric elements of $L(E)$, 145
$tr(A)$	trace of A.
A^*	adjoint of A.
$\mathcal{E}$	space of states, 146
ρ	density matrix, 147
$\mathcal{E}'$	space of density matrices, 147